Chapter 4
Optical Modulation and Coding

Samuel J. Dolinar, Jon Hamkins, Bruce E. Moision, and Victor A. Vilnrotter

4.1 Introduction

It can be argued that optical communications had its origins in ancient times, where modulated sunlight was often used to convey information over large distances. For example, mirrors have been used to create bright flashes of light in certain directions, producing a form of on–off modulation. Similarly, blankets used to cover a signal fire periodically produced puffs of smoke that could be seen for miles in the clear desert air, producing, in effect, a modulated signal. In modern times, navies throughout the world used bright incandescent light sources directed by a reflector and blanked by a manual shutter to send messages between ships often kilometers apart on the high seas. Before the invention of the laser, long-range optical communication was envisioned using bright flashes of light produced by intense pulses of electric current passing through an incandescent fiber placed in the focal-plane of an optical reflector. Even exploding wires that generated bursts of intense optical energy were considered for long-range applications. However, these sources were not effective in the production of intense, highly directional optical energy that could also be modulated at high enough data-rates to be seriously considered for deep-space optical communications.

The invention of the laser by Schawlow, Townes and Maiman [1,2] ushered in the era of deep-space optical communications. Here was a source of intense, highly directed optical energy that could produce coherent radiation, like radio frequency (RF) transmitters, but at much higher optical wavelengths. Because of its short wavelength the optical beam produced by a laser could be highly concentrated in the desired direction, constrained only by diffraction effects imposed by the finite extent of the transmitter aperture. Therefore, much higher concentrations of signal energy could be delivered to a distant receiver than with radio frequencies, suggesting the potential for communicating at higher data rates for a given amount of expanded power. No longer are optical communications all smoke and mirrors!

It was quickly realized that lasers enabled a new type of modulation format that was not practical at radio frequencies, namely, intensity modulation. The same phase-coherent modulation techniques developed for radio frequencies could still be employed with lasers, but the potential of intensity modulation promised additional choices. This new capability is a direct result of the relatively high energy of optical photons, enabling the detection of individual photons at optical wavelengths, which is much more difficult at radio frequencies. Optical frequencies stand at a nexus of the frequency spectrum where both wave and particle views of light are useful concepts.

Varying the intensity of the laser suggests the use of on–off modulation concepts reminiscent of earlier attempts with solar radiation and incandescent sources, but potentially at a much higher rate. Surprisingly, it has been found [3] that relying entirely on the energy in the signal and ignoring the coherence of the radiation does not fundamentally limit the rate at which information can be transferred reliably over the noiseless optical channel.

In the particle view of light, one can naively imagine using individual photons of light to carry information. For example, to communicate one bit of information, simply transmit one photon or not, depending on the value of the bit. This would be intensity modulation carried to its most granular extreme. If the information bits were equally likely 0 or 1, and if all photons were noiselessly detectable, this scheme would achieve an average photon efficiency of two bits per photon.

Optical photons do not behave exactly as classical particles, and correct quantum mechanical models are needed to describe their generation, modulation, and detection. However, the particle view of light is still a very useful concept for interpreting the efficiencies of optical modulations. Practical optical modulations can communicate information at reasonable fidelity with efficiencies ranging from a handful of bits per (detected) photon to a handful of (detected) photons per bit.

In this chapter, we discuss modulation and coding for the optical communications channel. For this purpose, the optical communications channel may be reduced to the block diagram of Fig. 4-1. User information, denoted U, is first encoded by an error correction code (ECC), mapping user bits to code bits C. The ECC introduces redundant information into the message to aid in correcting errors in the received signal. The code bits are passed though a modulator, which maps the coded bits to symbols X. The set of symbols represents the distinct messages the laser will transmit. For example, the symbols could be distinct phase, polarization, wavelength, or amplitudes of the optical carrier. These symbols are transmitted over the noisy optical channel, detected and received as noisy versions Y. The process of encoding and modulation is then inverted by the demodulator and decoder. The demodulator takes the noisy received signal and produces estimates of the transmitted symbols $\hat{X}$, or the code symbols $\hat{C}$, or both. The decoder operates on these estimates to yield estimates of the user information $\hat{U}$. In modern modulation and coding design, demodulation and decoding are executed iteratively, illustrated in Fig. 4-l by a pair of directed arrows, passing revised estimates back and forth until the two reach agreement.

In addition to the throughput of the optical communications link, measured in bits/second, we are interested in a measure of its efficiency–a measure of how well we are using the available resources to achieve a desired throughput. A commonly used measure is the photon efficiency ρ, the number of user bits transmitted per signal photon [3,4]. As in [7], the throughput, R, can be written as the product of the photon efficiency and the average signal photons received per second, n_s:

$$R = \rho n_s \quad \text{bits/second} \tag{4.1-1}$$

The usefulness of Eq. (4.1-1) stems from the fact that it separates the communications problem into two parts that can often be treated independently: one relating to the generation and transmission of photons, and one relating to modulation and coding. In cases such that ρ depends only on the modulation and coding and n_s depends only on the characteristics of the laser, the data-rate R can be maximized for a given photon rate by maximizing ρ, subject to constraints imposed by bandwidth, peak power, and other limitations of the receiving and transmitting equipment. This interpretation is valid for a large class

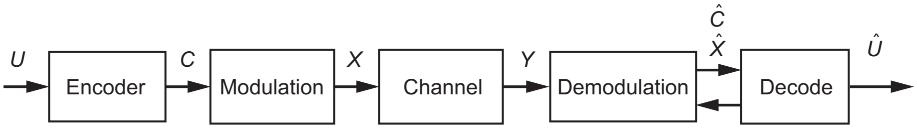

Fig. 4-1. The optimal communications channel.

of modulations, namely those that do not reduce the average photon rate at the receiver: examples include all lossless modulation formats such as phase, polarization, wavelength, and all rate-conserving intensity modulations. However, a decoupled model is often not valid. In practical intensity-modulated systems, the average amount of energy used by the laser-modulator may be a function of the timetable selected by the modulator for transmitting photons. This coupling of the characteristics of the laser and the modulator can destroy the separability of parameters expressed in Eq. (4.1-1). If it is found that while maximizing ρ a point is reached where further increases in ρ begin to reduce n_s, thus violating the independence assumption, then it becomes necessary to consider both ρ and n_s together when attempting to maximize R. In any case, it is clear that high rates of information transfer between transmitter and receiver require modulation and coding schemes capable of high information efficiencies. A thorough description of the generation, detection, and evaluation of efficient modulation and coding schemes capable of operating reliably with high information efficiencies is the subject of the rest of this chapter.

The sections of this chapter each discuss one or more components of Fig. 4-1, working roughly in order of dependence, starting with Section 4.2, which addresses the channel model. In Section 4.2, we discuss the quantum representation of optical fields and models for optical transmission and detection. This results in various statistical characterizations of a detected optical signal at the receiver. These models are the required starting point for the analysis and selection of appropriate modulation and coding, as well as determining the fundamental limits of communication with the optical channel.

Sections 4.3 and 4.4 address modulation. Section 4.3 discusses various common optical modulation schemes. Section 4.4 shows how these common optical modulations can be placed in a unified framework as cases of constrained on–off keying. This framework is used to introduce practical physical constraints on the modulation and analyze their impact. Section 4.5 addresses the performance of the demodulator, analyzing the performance of the uncoded modulation schemes on various channel models.

In Section 4.6 we determine channel capacity limits for the channel models developed in Section 4.2 when used with the modulation formats discussed in Section 4.3. The channel capacity yields bounds on the rate of information transmission as a function of the available physical resources subject to a fidelity criteria, e.g., the maximum bits per second achievable with a given average laser power and probability of error of 10^{-6}. We address physical constraints such as average power, peak power, bandwidth, and decision method at the detector (hard decision versus soft decision).

Section 4.7 discusses various types of channel codes that may be used to encode the user data. Finally, with all the pieces in place, in Section 4.8 we illustrate the performance of the coded modulation schemes on optical channel models, comparing performance to the theoretical capacity limits.

4.2 Statistical Models for the Detected Optical Field

In optical systems the statistical nature of the channel output depends on the detection method. At the receiver, light is focused onto a detector. There are two popular methods for detecting the received optical field. The first method, called *direct detection*, allows the received field to impinge directly upon the photodetector, which responds to its energy, i.e., to the squared magnitude of the incident optical field. Direct detection is non-coherent detection because any information in the phase of the received signal is lost. The second method, called *coherent detection*, adds a strong local optical field to the received field prior to photodetection.

In either case, the detector output may be either discrete or continuous. In most practical detectors—including photo-multiplier tubes (PMTs), avalanche photo-diode (APD) detectors, and PIN diode detectors—the output is a real-valued voltage or current that arises from the detector input as well as from random processes within the detector and follow-on circuitry.

In this section we give a brief formulation of the quantum optical field, and the resulting statistical models of the received optical signal when it is detected coherently or non-coherently.

4.2.1 Quantum Models of the Optical Field

Unlike familiar RF systems, optical systems often operate in regimes where the communication performance is susceptible to the fundamental measurement uncertainties implied by quantum theory. This observation has two ramifications for analyzing optical systems. First, mathematical models that accurately describe physically realizable optical fields and optical detectors must be firmly rooted in quantum theory. Second, the quantum theory permits additional abstract measurements formulated as mathematical operators in a Hilbert space that may in principle outperform the more conventional measurements that can be implemented by direct detection or coherent detection. Considerable attention in the literature [3,5] has been devoted to formulating and analyzing optical communications problems in abstract quantum Hilbert spaces.

The performance of some communications systems is ultimately limited by thermal noise entering the receiver along with the signal. This idea can be illustrated with Shannon's capacity formula for bandlimited classical channels, according to which error-free communication is possible at rates less than $W \log_2(1 + [S/N])$ bits per second (bps) [6], where W is the bandwidth in hertz, S is the average signal power, and N is the power of additive thermal noise. The above expression shows that if there is no thermal noise, then error-free communication should be possible at arbitrarily high rates, even with bandlimited channels. However, the Shannon formulation implicitly assumes that arbitrarily precise measurements are possible, since if this were not the case then there would have to be an effective noise term associated with the noisy signal measurements in the denominator of the capacity expression in addition to the thermal noise, and hence the denominator would never actually approach zero.

The classical model assumes that deterministic signals are observed in the presence of additive Gaussian noise. This model is perfectly adequate for describing communications systems operating at radio frequencies, where quantum effects are not readily detectable. However, at optical frequencies quantum effects tend to be the dominant source of error, and therefore must be considered in the communications system model. The approach most consistent with the principles of quantum mechanics starts out by quantizing the received electromagnetic field, and seeks to determine those measurements on the received field that achieve the best results such as, for example, minimizing the average probability of detection error. The best measurements may not be readily realizable with physically available devices; however, it is often possible to determine the performance of the "quantum optimum" receiver analytically. Therefore, if these measurements could somehow be made and incorporated into a communications receiver, the performance of the "quantum optimum" receiver would represent the achievable limit on communications system performance consistent with the principles of quantum mechanics.

Another approach for evaluating optical communications systems assumes a classical instead of a quantized received field, but models the response of physically realizable detectors using the same statistics that a quantum mechanical model would provide [7]. This "quantum mechanically correct" detector response is then used as the fundamental observable on which the decisions are based. Receivers using this approach are often called "semi-classical" and have the advantage of employing well-known detection techniques; however, such receivers generally cannot match the performance of the optimum quantum receiver.

4.2.1.1 Quantization of the Electric Field. The application of detection theory to a quantum mechanical model of the aperture field was originally developed by Helstrom and summarized in the IEEE Proceedings article "Quantum Mechanical Communications Theory" [5], as well as the subsequent monograph

Quantum Detection and Estimation Theory, Mathematics in Science and Engineering and the references therein [3]. This section relies heavily on results presented therein and on the numerous references cited in these publications. A detailed derivation of quantum communication theory is beyond the scope of this chapter; therefore, we concentrate on summarizing key results, with a minimum of discussion and derivation.

The receiver in a quantum optical communications system is often modeled as a large cubical box of volume V, with perfectly conducting walls [5]. The received field is admitted into this box, or cavity, through a receiving aperture assumed normal to the direction of propagation, during the time interval $(0,T)$. After this time the aperture is closed, and measurements are made on the "received field" inside the cavity. The received field can be represented as a superposition of normal modes of the cavity, where each mode behaves much like a harmonic oscillator with radian frequency ω_k. The classical waveform $\varepsilon(\mathbf{r},t)$ can be expanded in terms of standing-wave normal mode functions $u_k(\mathbf{r})$ within the cavity as [3,5]

$$\varepsilon(\mathbf{r},t) = -\varepsilon_0^{-1/2} \sum_k p_k(t)\, u_k(\mathbf{r}) \tag{4.2-1}$$

where ε_0 is the dielectric constant for free space, and $p_k(t)$ and $u_k(\mathbf{r})$ describe the temporal and spatial variation of the kth field mode, respectively. If the received field is represented in terms of plane-wave mode functions (instead of standing waves), the spatial variation of each mode takes the form

$$u_k(\mathbf{r}) = V^{-1/2}\, \mathbf{e}_k \exp(i\mathbf{k}\cdot\mathbf{r}) \tag{4.2-2}$$

where $\mathbf{k}$ is the propagation vector and $\mathbf{e}_k$ is a unit polarization vector perpendicular to $\mathbf{k}$. The complex amplitude $\alpha_k(t)$ of each traveling-wave mode can be expressed as

$$\alpha_k(t) = \alpha_k \exp(-\,i\,\omega_k t) \tag{4.2-3}$$

In the quantum formulation, the coordinates and momenta of the harmonic oscillators are replaced by the corresponding operators $Q_k(t)$ and $P_k(t)$, while the complex amplitudes are replaced by the operators a_k. The operator $a_k^+ a_k$ is called the "number operator" because its eigenvectors are the "number states" denoted by the ket $|n\,\rangle$, and its eigenvalues are the non-negative integers:

$$a_k^+ a_k |n_k\rangle = n_k |n_k\rangle \tag{4.2-4}$$

The operator a_k converts the eigenvector $|n_k\rangle$ to $|n_k - 1\rangle$, whereas the operator a_k^+converts $|n_k\rangle$ to $|n_k + 1\rangle$: for this reason, these operators are often called "annihilation" and "creation" operators.

The electric field operator can be expressed in terms of the annihilation and creation operators [3,5] as

$$E(\mathbf{r},t) = i \sum_k \frac{\hbar\omega_k}{\sqrt{2\varepsilon_0 V}} \mathbf{e}_k \left\{ a_k \exp[-i\,(\omega_k t - \mathbf{k}\cdot\mathbf{r}] - a_k^+ \exp[i\,(\omega_k t - \mathbf{k}\cdot\mathbf{r})\right\} \tag{4.2-5}$$

4.2.1.2 The Coherent State Representation of a Single Field Mode.

When the k-th mode of the electric field is in a state that is the right eigenvector of the annihilation operator, $a_k|\alpha_k\rangle = \alpha_k|\alpha_k\rangle$, it is said to be in a coherent state. The coherent states are denoted by kets $|\alpha_k\rangle$, and can be expressed in terms of the number eigenstates as [8,9]

$$|\alpha_k\rangle = \exp[-|\alpha_k|^2/2] \sum_{n_k=0}^{\infty} (n_k!)^{-1/2}\, \alpha_k^{n_k} |n_k\rangle \tag{4.2-6}$$

The coherent states are normalized so that $\langle \alpha_k|\alpha_k\rangle = 1$. The overlap between two coherent states, $|\alpha_k\rangle$ and $|\beta_k\rangle$ is not zero; hence, the coherent states are not orthogonal. Denoting the average number of photons in the k-th normal mode by $K_{sk} \equiv |\alpha_k|^2$, the probability that n photons are contained in that mode can be found as

$$\Pr(n) = |\langle n|\alpha_k\rangle|^2 = K_{sk}^n \exp[-K_{sk}]/n! \tag{4.2-7}$$

which is recognized as the Poisson probability mass function (pmf).

Although the coherent states are not orthogonal, they are complete, and hence can be used to expand a large class of density operators, including those of interest in communication theory, as follows [3,5,9]:

$$\rho = \int P(\alpha_k) \prod_k |\alpha_k\rangle \langle \alpha_k|\, d^2\alpha_k \tag{4.2-8}$$

An expansion of this form is called the P representation, where $P(\{\alpha_k\})$is called the weight function. The P representation will be used to describe the density operators of fields consisting of thermal noise, and signal plus thermal noise, in the following development.

4.2.1.3 Quantum Representation of Thermal Noise. When the field inside the cavity consists of thermal radiation alone, in equilibrium with the cavity at an absolute temperature of T_{th} kelvins, the density operator describing the state of the k-th normal mode can be expressed in terms of the P representation as [3,5,9]

$$\rho_k = (\pi K_k)^{-1/2} \int \exp[-|\alpha|^2/K_k]|\alpha\rangle\langle\alpha|d^2\alpha \tag{4.2-9}$$

where K_k is the average number of photons in the k-th mode. Making use of the number-state expansion of the coherent states in Eq. (4.2-6), this density operator can also be expressed in terms of the number states as

$$\rho_k = \sum_{n_k=0}^{\infty} (1 - v_k)\, v_k^{n_k}\, |n_k\,\rangle\,\langle\, n_k| \tag{4.2-10}$$

where $v_k = K_k/(K_k + 1)$. Note that this density operator is diagonal in the number representation. The probability of finding a given number of photons in the kth mode is simply

$$\Pr(n) = \langle\, n|\rho|n\,\rangle = \sum_{n_k=0}^{\infty} (1 - v_k)\, v_k^{n_k}\, \langle\, n|n_k\,\rangle\,\langle\, n_k|\, n\,\rangle = (1 - v_k)\, v_k^{n} \tag{4.2-11}$$

which is recognized as a Bose–Einstein probability. In the classical limit, when the inequality $\kappa T \gg \hbar\,\omega_k$ holds (here κ is Boltzmann's constant), the weight function of the k-th mode becomes

$$P(\alpha_k) = (\pi K_k)^{-1/2} \exp[-|\alpha_k|^2/K_k] \tag{4.2-12}$$

This is recognized as a Gaussian probability density, representing the probability density of the complex envelope of the kth normal mode.

4.2.1.4 Quantum Representation of Signal Plus Thermal Noise. When the received field mode contains both thermal noise and a coherent-state signal component $|\mu_k\,\rangle$, the density operator is of the form

$$\rho_k = (\pi K_k)^{-1/2} \int \exp[-|\alpha - \mu_k|^2/K_k]|\alpha\rangle\langle\alpha|d^2\alpha \tag{4.2-13}$$

This density operator can also be expressed in terms of number states, as shown in [3,5]. With the number state representation, this density operator can be interpreted as an infinite-dimensional matrix with elements

$$\langle n|\rho_k|m\rangle = (1 - v_k)\sqrt{\frac{n!}{m!}} v_k^m \left(\frac{\mu_k^*}{K_k}\right)^{m-n} e^{-(1-v_k)|\mu_k|^2} L_n^{m-n}\left[\frac{-(1-v_k)^2|\mu_k|^2}{v_k}\right] \tag{4.2-14}$$

where $L_n^{m-n}(x)$ is the generalized Laguerre polynomial. If no attempt is made to maintain coherence between the transmitter and the receiver, the density operator must be averaged with respect to phase. If the phase is taken to be uniformly distributed between $(0, 2\pi)$, the averaged density operator $\bar{\rho}_k$ becomes diagonal in the number representation:

$$\bar{\rho}_k = \sum_{k=0}^{\infty} (1 - v_k)\, v_k^n \exp[-\,(1 - v_k)K_{sk}]\, L_n[-(1 - v_k)^2 K_{sk}/v_k]\, |n_k\,\rangle\,\langle\, n_k| \tag{4.2-15}$$

where $L_n(x)$ is the ordinary Laguerre polynomial, and $K_{sk} \equiv |\mu_k|^2$ is the average number of signal photons in the normal mode. As before, the probability of obtaining exactly n photons can be found as

$$\Pr(n) = \langle\, n|\bar{\rho}_k|n\,\rangle = (1 - v_k)\, v_k^n \exp[-\,(1 - v_k)K_{sk}]\, L_n[-(1 - v_k)^2 K_{sk}/v_k] \tag{4.2-16}$$

This is recognized as specifying the well-known Laguerre probabilities for the number of photons in a mode containing both signal fields and noise fields of thermal origin.

4.2.2 Statistical Models for Direct Detection

From a communications systems perspective, it is often convenient to state the various stochastic properties for each detection method in terms of a common framework. For many of the modulations and classical detection methods considered in this chapter, this may be done with the general binary-input discrete time channel model shown in Fig. 4-2. This model can be used for a direct

4.2.2.1.2 Limiting Form of Thermal Noise Fields. Thermal noise fields, and combinations of coherent signal and thermal noise fields, also generate approximate Poisson-distributed random variables at the output of an ideal photon counter, under certain limiting conditions. Whereas the number of photons, hence the modal count, for a single mode of thermal radiation is Bose–Einstein distributed as shown in Eq. (4.2-11), the count distribution for thermal light filtered by a narrowband rectangular optical filter of bandwidth W Hz around the optical carrier frequency, and observed for T seconds, is given by the negative binomial density [7]

$$\Pr(k) = \binom{k + WT - 1}{WT - 1}(1 - v)^{WT}\, v^k \tag{4.2-19}$$

where $v = N/(N+1)$, and N is the average number of photons in a single mode of the thermal field. If the time-bandwidth product approaches infinity, $W\,T \to \infty$, and at the same time N decreases such that NWT remains constant, the distribution for the total number of photons approaches the Poisson limit:

$$f_{Y|X}(k|x=0) = \lim_{\substack{WT\to\infty \\ NWT=c}} \Pr(k) = \frac{K_b^k e^{-K_b}}{k!} \tag{4.2-20}$$

where K_b is the sum of the average numbers of modal background photon counts. This can be extended to the case of signal plus thermal noise: K_b is replaced by $K_s + K_b$ in Eq. (4.2-20) when $x = 1$, as shown in [7].

4.2.2.2 The McIntyre–Conradi Model for APD Detectors. The average number of photons absorbed over the active surface of an APD illuminated with optical power $P(t)$ in T seconds can be expressed as [32]

$$K = \frac{\eta}{h\nu}\int_0^T P(t)\,dt \tag{4.2-21}$$

where h is Planck's constant, ν is the optical frequency, and η is the detector's quantum efficiency, defined as the average number of photons absorbed by the APD's photosensitive surface divided by the average number of incident photons. The actual number of absorbed photons is a Poisson-distributed random variable with mean K (where $K = K_b$ or $K = K_s + K_b$, as in Eqs. (4.2-17) and (4.2-18)).

In an APD, the density of the output electrons in response to absorbed photons was modeled accurately by McIntyre [40] and verified experimentally

4.2.2.1.2 Limiting Form of Thermal Noise Fields. Thermal noise fields, and combinations of coherent signal and thermal noise fields, also generate approximate Poisson-distributed random variables at the output of an ideal photon counter, under certain limiting conditions. Whereas the number of photons, hence the modal count, for a single mode of thermal radiation is Bose–Einstein distributed as shown in Eq. (4.2-11), the count distribution for thermal light filtered by a narrowband rectangular optical filter of bandwidth W Hz around the optical carrier frequency, and observed for T seconds, is given by the generalized Bose–Einstein density [7]

$$\Pr(k) = \binom{k + WT - 1}{WT - 1} (1 - v)^{WT} \, v^k \tag{4.2-19}$$

where $v = K/(K + 1)$, and K is the average number of photons in a single mode of the thermal field. If the time-bandwidth product approaches infinity, $W\,T \to \infty$, and at the same time K decreases such that KWT remains constant, the distribution for the total number of photons approaches the Poisson limit:

$$f_{Y|X}(k|x) = \lim_{\substack{WT\to\infty \\ KWT=c}} \Pr(k) = \frac{K^k e^{-K}}{k!} \tag{4.2-20}$$

where c is a constant and K is the sum of the average number of modal photon counts, i.e., $K = K_b$ when $x = 0$ and $K = K_s + K_b$ when $x = 1$.

4.2.2.2 The McIntyre–Conradi Model for APD Detectors. The average number of photons absorbed over the active surface of an APD illuminated with optical power $P(t)$ in T seconds can be expressed as [32]

$$K = \frac{\eta}{h\nu} \int_0^T P(t)\, dt \tag{4.2-21}$$

where h is Planck's constant, ν is the optical frequency, and η is the detector's quantum efficiency, defined as the average number of photons absorbed by the APD's photosensitive surface divided by the average number of incident photons. The actual number of absorbed photons, call it N, is a Poisson-distributed random variable with mean K (where $K = K_b$ or $K = K_s + K_b$, as in Eqs. (4.2-17) and (4.2-18)).

In an APD, the density of the output electrons in response to absorbed photons was modeled accurately by McIntyre [40] and verified experimentally by Conradi [40]. We refer to it as the McIntyre–Conradi distribution. The conditional density of obtaining k electrons at the APD output in response to $N = n$ absorbed photons is given by

$$f_{Y|N}(k|n) = \frac{n\Gamma\left(\dfrac{k}{1-k_{eff}}+1\right)}{k(k-n)!\Gamma\left(\dfrac{k_{eff}k}{1-k_{eff}}+n+1\right)}\left[\frac{1+k_{eff}(G-1)}{G}\right]^{\frac{n+k_{eff}k}{1-k_{eff}}}$$

$$\times\left[\frac{(1-k_{eff})(G-1)}{G}\right]^{k-n} \tag{4.2-22}$$

where G is the average gain of the APD, and k_{eff} is the ionization ratio, $0 < k_{eff} < 1$, a property of the semiconductor. Averaging Eq. (4.2-22) over the number of absorbed photons, n, yields

$$f_Y(k) = \sum_{n=1}^{k} f_{Y|N}(k|n)\,\frac{K^n}{n!}\,e^{-K}, \quad k \geq 1 \tag{4.2-23}$$

where the summation limit is k instead of infinity because according to the model there can never be more absorbed photons than released electrons. Thus, for $k \in \mathbb{N}$ (where $\mathbb{N}$ is the set of natural numbers),

$$f_{Y|X}(k|x) = \sum_{n=1}^{k}\frac{n\Gamma\left(\dfrac{k}{1-k_{eff}}+1\right)\left[\dfrac{1+k_{eff}(G-1)}{G}\right]^{\frac{n+k_{eff}k}{1-k_{eff}}}K_x^n e^{-K_x}}{k(k-n)!\Gamma\left(\dfrac{k_{eff}k}{1-k_{eff}}+n+1\right)n!}$$

$$\times\left[\frac{(1-k_{eff})(G-1)}{G}\right]^{k-n} \tag{4.2-24}$$

where $K_0 = K_b$ is the average number of photons detected when $x = 0$ and $K_1 = K_s + K_b$ is the average number of photons detected when $x = 1$. In determining K for an APD, there are some subtleties to Eq. (4.2-21) that should be noted. The APD bulk leakage current, I_b, is multiplied by the APD gain and can be modeled artificially as part of the background optical power $P(t)$ entering the telescope, even when no actual background light is present. The APD surface leakage current, I_s, is not multiplied by the APD gain and can be modeled as a direct current (DC) at the output.

4.2.2.3 The Webb, McIntyre, and Conradi Approximation to the McIntyre–Conradi Model. An approximation to Eq. (4.2-24) that is simpler to evaluate has been derived by Webb, McIntyre, and Conradi [41], which we refer to as the WMC density, and is given by the continuous conditional density function

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi KG^2F}}\left(1+\frac{(y-GK)(F-1)}{KGF}\right)^{-3/2} \times \exp\left(\frac{-(y-GK)^2}{2KG^2F\left(1+\dfrac{(y-GK)(F-1)}{KGF}\right)}\right) \tag{4.2-25}$$

for $y \geq K$, where the excess noise factor, F, is defined as $F = k_{\mathit{eff}}G + (2-[1/G])\,(1-k_{\mathit{eff}})$. Unfortunately, the constraint $y \geq K$ causes this to be an invalid density function–it does not integrate to one. If we extend the domain artificially, this difficulty is avoided. Denoting the zero-mean, unit-variance WMC probability density function (pdf) (with valid domain indicated) by

$$p_w(y;\delta) = \frac{1}{\sqrt{2\pi}}(1+y/\delta)^{-3/2}\exp\left[\frac{-y^2}{2(1+y/\delta)}\right], \quad y > -\delta \tag{4.2-26}$$

the conditional density at the output of an APD can be written as

$$f_{Y|X}(y|x) = \frac{1}{\sigma_x}p_w\left(\frac{y-m_x}{\sigma_x};\delta_x\right) \tag{4.2-27}$$

where $x \in \{0,1\}$, $m_x = K_xG$, $\sigma_x^2 = K_xG^2F$, $\delta_x^2 = K_xF/(F-1)^2$, $K_0 = K_b$, and $K_1 = K_s + K_b$.

Unlike a Gaussian distribution, the WMC distribution is not determined solely by its mean and variance; it also depends on the "skewness" parameter δ. As $\delta \to \infty$, the WMC distribution reduces to the normal distribution, as illustrated in Fig. 4-3. However, the model does not allow independently varying skewness and variance parameters. Both the skewness and variance are signal dependent, being proportional to the average number of absorbed photons, K_x. Thus, this model imposes the constraint $\sigma_0/\sigma_1 = \delta_0/\delta_1$. We shall use this constraint for the WMC channel in the remainder of this chapter.

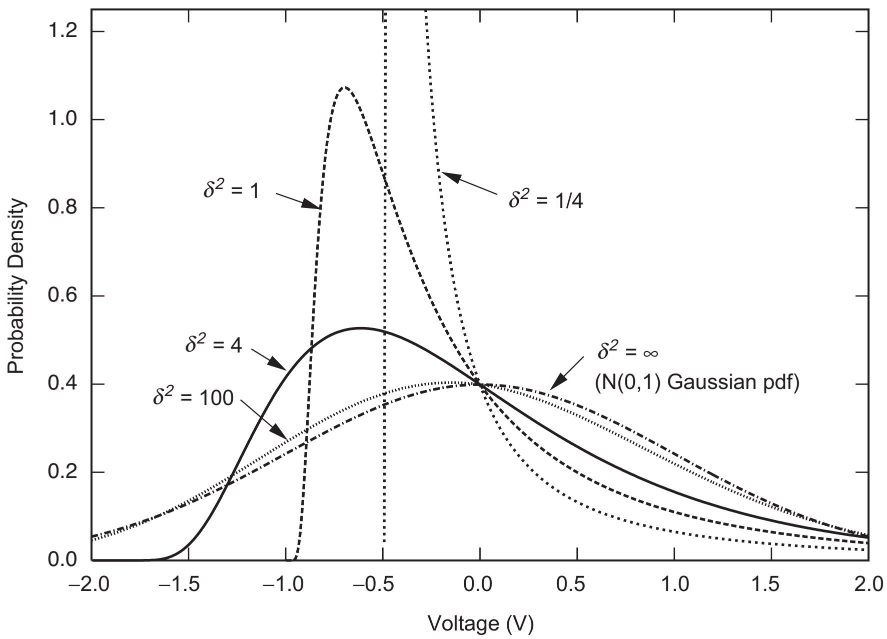

Fig. 4-3. The standardized WMC pdf, for various δ.

4.2.2.4 The WMC Plus Gaussian Approximation. In addition to the avalanche electrons produced by the APD, the follow-on amplifier and resistance within the follow-on electronics also generate electrons of thermal origin, which can be taken into account [42]. That is, the detector output Y is the sum of the McIntyre–Conradi component, with the conditional density function given by Eq. (4.2-24) and an independent normal (Gaussian) component with zero mean and variance σ_n^2. If we use the WMC density in Eq. (4.2-27) to approximate Eq. (4.2-24), the WMC plus Gaussian conditional probability density function becomes

$$f_{Y|X}(y|x) = \int_{m_x-\delta_x\sigma_x}^{\infty} \underbrace{\frac{1}{\sigma_n}\phi\left(\frac{y-z}{\sigma_n}\right)}_{\text{Gaussian pdf at } y-z} \cdot \underbrace{\frac{1}{\sigma_x}p_w\left(\frac{z-m_x}{\sigma_x};\delta_x\right)}_{\text{WMC pdf at } z} dz$$

where again $x \in \{0,1\}$ indicates the condition that either a 0 or a 1 was sent, and $\phi(x)$ is the normal density, $\phi(x) = 1/\sqrt{2\pi}e^{-x^2/2}$.

4.2.2.5 Additive White Gaussian Noise Approximation. Under certain conditions, the additive white Gaussian noise (AWGN) channel model can be used to model direct detection [42], and it is a model that has been used in free-space optical communications link budget software at NASA [34]. The conditional pdf at the output of the channel is given by

$$f_{Y|X}(y|x) = \frac{1}{\sigma_x}\phi\left(\frac{y-m_x}{\sigma_x}\right) = \frac{1}{\sqrt{2\pi\sigma_x^2}}e^{-(x-m_x)^2/2\sigma_x^2} \tag{4.2-28}$$

where $x \in \{0,1\}$, and m_x and σ_x^2 are the mean and variance, respectively, when $X = x$. This model is often only an approximation, and it may over- or underestimate error probabilities [49,50]. In this subsection, we discuss several ways in which this model may arise.

4.2.2.5.1 AWGN Approximation for an APD. The probability density of Y is approximately Gaussian when operating under high background conditions, and in the presence of additive Gaussian noise [49–51]. In this case, the conditional density is given by Eq. (4.2-28), with [51]

$$m_0 = G\,K_b + I_sT_s/q \tag{4.2-29}$$

$$m_1 = G(K_s + K_b) + I_sT_s/q \tag{4.2-30}$$

$$\sigma_0^2 = \left[G^2FK_b + \frac{I_sT_s}{q} + \frac{2\kappa TT_s}{q^2R_L}\right]2BT_s \tag{4.2-31}$$

$$\sigma_1^2 = \left[G^2F(K_s + K_b) + \frac{I_sT_s}{q} + \frac{2\kappa TT_s}{q^2R_L}\right]2BT_s \tag{4.2-32}$$

where G is the average APD gain, K_b and K_s are the average number of absorbed background and signal photons, respectively, I_s is the surface leakage current of the APD, T_s is the slot width, and q is the electron charge. The other parameters are the "excess noise" factor of the APD F, the equivalent noise temperature of the device T, the Boltzmann's constant κ. The noise bandwidth B is assumed to be matched to the slot duration as $B = 1/2T_s$. As mentioned earlier, the bulk dark current of the APD, I_b, can be artificially modeled as part of the background radiation, by absorbing the quantity I_b/q into K_b.

4.2.2.5.2 AWGN Approximation for the Ideal PMT. The conditional probability density of the output voltage Y, given X, of a PMT can be expressed as the sum of conditional densities representing the output voltage for a given number of absorbed photons, n:

$$f_{Y|X}(y|x) = \sum_{n=0}^{\infty} f_{Y|N}(y|n) f_{N|X}(n|x) \tag{4.2-33}$$

where $f_{N|X}(n|x)$ is a Poisson-probability mass function with mean K_b when $x = 0$ and mean $K_s + K_b$ when $x = 1$. The conditional density of the output voltage may be modeled as Gaussian [32]:

$$f_{Y|N}(y|n) = \phi\left(\frac{y - Gn}{\sigma}\right) \tag{4.2-34}$$

where G is the average gain of the PMT and $\sigma = \xi G n$, where ξ is the spreading factor of the PMT [32].

During daytime operation, or with a bright planet (Mars) in the field of view, the Poisson-distributed number of absorbed photons, N, can be much greater than one, which justifies a continuous approximation. A Gaussian approximation of $f_{N|X}(n|x)$ may be made [52], making Eq. (4.2-33) a sum of Gaussians, and hence Gaussian as well:

$$f_{Y|X}(y|x) = \phi\left(\frac{y - G\bar{N}_x}{\sigma_x}\right) \tag{4.2-35}$$

where $\sigma_x = \xi G \bar{N}_x$, with $N_x = K_b$ when $x = 0$ and $N_x = K_s + K_b$ when $x = 1$.

4.2.3 Summary of Statistical Models

Table 4-1 lists the conditional density (or mass) function $f_{Y|X}(y|x)$ associated with each model discussed earlier in this section. In each case, m_0 and σ_0^2 denote the mean and variance conditioned on $X = 0$ (see Fig. 4-2), and m_1 and σ_1^2 denote the corresponding quantities conditioned on $X = 1$. We define the slot signal-to-noise ratio (SNR) as $\beta = (m_1 - m_0)^2/\sigma_0^2$, the "excess SNR" as $\gamma = (m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$, and the "bit SNR" as $\beta_b = \beta/(2R_c)$, where R_c is the number of bits per symbol.

4.3 Modulation Formats

Optical direct detection effectively measures the energy in the optical signal impinging on the detector. Since direct detection does not respond to phase, but is capable of distinguishing only between different intensity levels, intensity modulation is required. The most common forms of digital intensity

Table 4-1. The pmf's and pdf's of several optical channel models. In each case, for $x \in \{0,1\}$, m_x and σ_x^2, respectively, are the mean and variance when $X = x$.

Channel Model	pmf or pdf
Poisson	$f_{Y\|X}(k\|x) = \frac{m_x^k e^{-m_x}}{k!}$, where $m_0 = K_b, m_1 = K_s + K_b, k \in \mathbb{N}$
AWGN	$f_{Y\|X}(y\|x) = \frac{1}{\sigma_x}\phi\left(\frac{y-m_x}{\sigma_x}\right) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{(y-m_x)^2}{2\sigma_x^2}\right], y \in \mathbb{R}$ (where $\mathbb{R}$ is the set of real numbers)
McIntyre-Conradi	$f_{Y\|X}(k\|x) = \sum_{n=1}^{k} \frac{n\Gamma\left(\frac{k}{1-k_{eff}}+1\right)\left[\frac{1+k_{eff}(G-1)}{G}\right]^{n+k_{eff}k/(1-k_{eff})}\left[\frac{(1-k_{eff})(G-1)}{G}\right]^{k-n} K_x^n e^{-K_x}}{k(k-n)!\Gamma\left(\frac{k_{eff}k}{1-k_{eff}}+n+1\right)n!}$
WMC	$f_{Y\|X}(y\|x) = \frac{1}{\sqrt{2\pi\sigma_x^2}}\left(1+\frac{y-m_x}{\sigma_x\delta_x}\right)^{-3/2} \exp\left[-\frac{(y-m_x)^2}{2\sigma_x^2\left(1+\frac{y-m_x}{\sigma_x\delta_x}\right)}\right], y \in \mathbb{R}$
WMC+Gaussian	$f_{Y\|X}(y\|x) = \int_{m_x-\delta_x\sigma_x}^{\infty} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{(y-t)^2}{2\sigma_n^2}\right] \cdot \frac{1}{\sqrt{2\pi\sigma_x^2}}\left(1+\frac{t-m_x}{\sigma_x\delta_x}\right)^{-3/2} \exp\left[-\frac{(y-m_x)^2}{2\sigma_x^2\left(1+\frac{t-m_x}{\sigma_x\delta_x}\right)}\right] dt$

modulation are on–off keying (OOK) and pulse-position modulation (PPM). In addition, wavelength modulation (WM) and pulse intensity modulation (PIM) can also be envisioned to increase the information throughput, but have not achieved the level of acceptance enjoyed by OOK and PPM. Wavelength modulation is a form of spatial PPM, hence its performance can be determined directly from PPM detection. Therefore, we shall concentrate on these modulation formats here, and only briefly describe other schemes.

Most of the common modulation formats suitable for direct detection can be decomposed into a simple binary on–off "pulsed" form of modulation. At a given wavelength or during a given period of time an optical pulse is either present or absent. The simplest of these is binary OOK, for which each modulated bit consists of one of these on–off pulses. More complex pulsed modulations, such as PPM, can be conceptually regarded as coded versions of binary OOK, as discussed in Section 4.4.

Coherent reception can be used with any pulsed modulation such as PPM, but additionally there are many non-pulsed modulations suitable for coherent reception. Examples include M-ary phase shift keying (M-PSK) and M-ary quadrature amplitude modulation (M-QAM), i.e., the whole suite of modulations used for coherent RF communication.

Pulsed modulations, such as OOK and PPM, share a common characteristic that the laser is either "on" or "off" during every slot time interval. Similarly, for wavelength-shift keying (WSK) laser energy is either present or absent in each frequency slot. Any such modulation can be conveniently regarded as a mapping from a binary sequence of 0's and 1's into a discrete set of optical pulses. A pulse is present in the ith slot if the ith bit is a 1, and the pulse is absent if the ith bit is a 0.

Pulsed modulations can be decomposed into a coded sequence of binary OOK pulses. Each pulsed modulation format, such as PPM, differential pulse-position modulation (DPPM), etc., simply imposes a different set of constraints on the binary coded sequence. For example, the coded sequences for M-ary PPM are the M binary sequences of length M containing exactly one 1 bit and $M-1$ 0 bits.

4.3.1 On–Off Keying (OOK)

With on–off keying, binary data is represented by the presence or absence of a single light pulse in each T-second symbol interval. The binary information sequence can be mapped directly to a sequence of light pulses at the transmitter according to the rule: if the information bit is 1, transmit a laser pulse; if it is 0, transmit nothing. Therefore, there is a one-to-one correspondence between 1's in the data-stream, and the occurrence of light pulses emanating from the transmitter. This is illustrated in Fig. 4-4. For comparison with other modulation schemes that use more than one slot per bit, alternate symbols are shaded.

4.3.2 Pulse-Position Modulation (PPM)

Optical PPM is well suited to existing laser modulation techniques (such as Q-switching, mode-locking, and cavity-dumping), requires low average power, attains reasonably high information efficiencies, and is resistant to background radiation.

A k-bit source $\mathbf{U} = (U_1, \cdots, U_k) \in \{0,1\}^k$ is modulated with M-ary PPM, $M = 2^k$, to yield a signal $\mathbf{X} = (0, \ldots, 0, 1, 0, \ldots, 0) \in \{0,1\}^M$, which contains a single one in the position indicated by the binary representation of $\mathbf{U}$. The transmission channel is a binary-input unconstrained-output memoryless channel ($X = 0$ or 1 in Fig. 4-5). One use of the overall PPM-symbol channel consists of M serial uses of the binary-input channel, and it produces the received vector $\mathbf{Y} = (Y_1, \cdots Y_M) \in \mathbb{R}^M$ This is illustrated in Fig. 4-6.

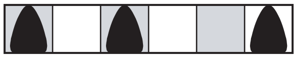

Fig. 4-4. OOK uses one slot per bit. Modulation of message 101001 is shown.

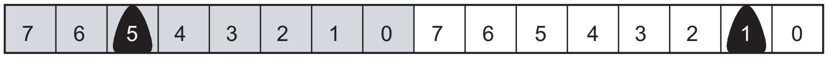
7
6
5
4
3
2
1
0
7
6
5
4
3
2
1
0

Fig. 4-5. 8-PPM uses an eight slot symbol for each three bits. Modulation of message 101001 is shown. The order in which slots are labeled is not consequential; here, label 7 is assigned to the left-most slot of a symbol, and label 0 is assigned to the right-most slot.

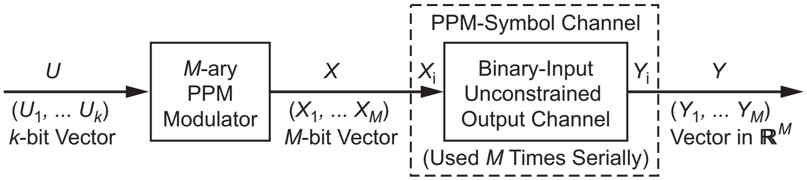
PPM-Symbol Channel
U
(U1, ... Uk)
k-bit Vector
M-ary
PPM
Modulator
X
(X1, ... XM)
M-bit Vector
Xi
Binary-Input
Unconstrained
Output Channel
Yi
(Used M Times Serially)
Y
(Y1, ... YM)
Vector in ℝM

Fig. 4-6. PPM signaling.

This modulation format consists of a fixed number of symbols, M, of equal duration, T seconds. Assuming for convenience that M is a power of two, $M = 2^L$, and recalling that the number of bits of information contained in a PPM symbol is $\log_2 M$, we can view the mapping from information-bits to PPM symbols as a one-to-one assignment of symbols to each of L consecutive information bits. To illustrate, consider the following sequences of $L = 4$ information bits and a particular mapping of a PPM pulse into one of $M = 2^L$ possible time-slots. Suppose each consecutive L-bit information sequence is mapped into a unique PPM symbol according to the rule, "1 plus the numerical value of the sequence, when the sequence is viewed as an L-digit binary number." According to this mapping, the sequence [0,0,0,0] is mapped to the first PPM slot at the transmitter since the numerical value of this sequence is 0. At the receiver, after slot and symbol synchronization has been achieved, this transmitted laser pulse gives rise to an average signal energy of K_s photons. Similarly, the sequence [1,0,0,1] corresponds to a laser pulse in the 10th slot, whereas the sequence [1,1,1,1] is mapped into the PPM symbol with a single pulse in the last, or 16th, time slot.

4.3.3 Differential PPM (DPPM)

In differential PPM [28], also called truncated PPM (TPPM), throughput is increased by beginning a new PPM symbol immediately following the slot containing the pulse. That is, non-pulsed slots of a PPM symbol which follow a pulsed slot are flushed. This imposes a more challenging synchronization problem because the symbols vary in length and error propagation could occur at the receiver or decoder. However, it also increases the throughput per unit time by a factor of two, since symbols are on average half as long as they would be with ordinary PPM. DPPM is shown in Fig. 4-7.

A Q-switched laser works well with the PPM format [53,54] because it can successfully confine a large pulse energy to a narrow slot. One side effect of Q-switched lasers, however, is a required delay, or dead time, between pulses during which the laser is recharged. This delay is significant relative to the pulse duration. PPM may be modified to satisfy the dead-time constraint by following each frame by a period during which no pulses are transmitted. However, this affects the optimality of PPM as a modulation format.

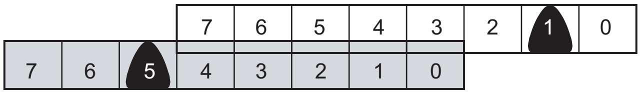

Fig. 4-7. 8-DPPM is the same as 8-PPM, except all slots following the pulse are flushed and a new symbol immediately follows. Modulation of message 101001 is shown.

DPPM is an attractive low-complexity, high-throughput scheme. However, it has two implementation issues that are common to variable-rate schemes. The first is the difficulty of adapting decoding algorithms to function with variable-rate codes. It may be particularly difficult to accommodate the code as an inner code in a concatenated coding scheme since a block of data from the outer code would map to a variable length block of transmitted symbols.

The second problem is the possibility of catastrophic error propagation in decoding due to loss of frame synchronization. Assume the DPPM decoder operates in a manner similar to a PPM decoder, by choosing the maximum slot count in an appropriate window. If an error is made in the estimation of the pulse position, the location of the following window will be incorrect. The detector will, however, re-synchronize with the next correctly detected pulse position. The probability of re-synchronizing in the frame following a pulse-position estimation is $\approx 2/3$ for q small and M large.

There may be methods of averting the problems with TPPM by buffering data and performing an appropriate sequence-detection algorithm. However, given lower-complexity options with similar performance, we did not pursue implementing DPPM.

4.3.4 Overlapping PPM (OPPM)

Overlapping PPM is a generalization of PPM proposed in [55]. In OPPM, each symbol interval of length T is divided into NM chips of duration $T_c = T/(NM)$. A pulse occupies N chips, and is constrained to be entirely contained within the symbol epoch. When $N = 1$, we have ordinary PPM discussed above, in which $\log_2 M$ bits are transmitted per T seconds. When $N > 1$, the pulse can be in one of $NM - N + 1$ positions, and we have $\log_2[NM - N + 1]$ bits per T seconds, i.e., nearly an additional $\log_2 N$ bits per T seconds for large M. OPPM signaling is shown in Fig. 4-8. OPPM imposes more stringent synchronization requirements, and special synchronizable codes may used to aid in this [29].

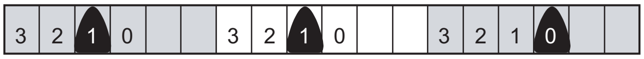

Fig. 4-8. OPPM with N = 3 and M = 2 has six chips per two bits. The four possible starting positions for three-chip-wide pulses are indicated. Modulation of message 101001 is shown.

4.3.5 Wavelength Shift Keying (WSK)

From the viewpoint of communication theory, wavelength-shift keying is similar to PPM of the same dimension. Instead of placing a single pulse of laser light of a given wavelength into one of M time-disjoint time-slots, WSK places a single laser pulse of one of M disjoint wavelengths into one of M optical detectors, as shown in Fig. 4-9.

4.3.6 Combined PPM and WSK

One way to avoid the linear increase in required average laser power with data-rate that is characteristic of WSK, is to use a combination of PPM and WSK, as shown in Fig. 4-10. By restricting the laser pulse to one of M time-slots, but allowing the laser pulse to take on any of N wavelengths, the low average laser power of conventional PPM can be maintained with the added advantage of increased data-rate. This can be demonstrated by observing that the dimensionality of the signal-space has been increased from M-dimensions (PPM) to NM dimensions.

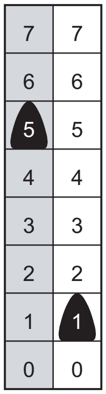

Fig. 4-9. 8-ary WSK uses eight frequencies in one time slot for each three bits. Modulation of message 101001 is shown.

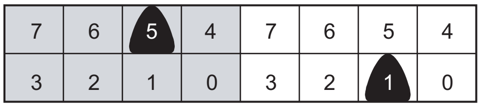

Fig. 4-10. Binary WSK with 4-PPM uses two frequencies and four time slots for each three bits. Modulation of message 101001 is shown.

4.4 Rate Limits Imposed by Constraints on Modulation

A modulation format enforces constraints on the relative location of pulses and describes a mapping of information bits to the sequence of pulses. The constraints are driven by the limitations of the physical devices and follow-on electronics. For example, Q-switched lasers require a minimum delay between pulses to allow the laser to recharge, while timing synchronization algorithms benefit from limiting the maximum delay between pulses. The modulation format is designed to satisfy such constraints while maximizing the throughput, or transmitted bits per second. The upper bound on the throughput for a set of constraints is referred to as the *Shannon capacity* of the constraint, to distinguish it from the channel capacity. The Shannon capacity may be thought of as the capacity of a channel with constraints on the inputs but no distortion from the channel. Given a set of practical constraints, we would like to determine the Shannon capacity and find modulations that satisfy the constraint in an efficient manner, i.e., with throughputs close to the Shannon capacity. This section treats those two problems.

If we represent a pulsed slot by a binary 1 and a non-pulsed slot by binary 0, the modulation format is seen to be a binary code mapping unconstrained binary sequences to constrained binary sequences. For example, PPM enforces the constraint that there is exactly one pulse within each frame of M slots and maps $\log_2 M$ bits to each of the M possible frames. It is, in essence, a binary code of rate $(\log_2 M)/M$. To reflect this, we refer to the modulation as a *modulation code*, a code designed primarily to satisfy certain signalling constraints, as opposed to correcting errors, although the distinction is muddled when both are involved in error correction, or are treated as a single code, as with some iterative decoding techniques. Any constraints on the sequence of 0's and 1's imposed by the physical devices and implemented by some modulation limits the amount of information per slot that can be conveyed. An unconstrained channel can support 1 bit/slot. In comparison, M-ary PPM conveys only $(\log_2 M)/M$ bits/slot, which decreases monotonically in M, for $M > 3$.

The study of Shannon capacity and modulation codes has taken on additional significance recently with the advent of new techniques in quantum mechanics capable of generating orthogonal quantum states. As we shall show in Section 4.5.4, when optimum quantum detection is employed, the probability of error for both binary and higher dimensional modulation depends on the overlap between the quantum states that represent the various hypotheses. The overlap approaches zero as the states become orthogonal, yielding error probabilities that also approach zero. Hence, error-free detection is possible when orthogonal quantum states are used. For this class of signals, the relevant channel capacity is the Shannon capacity.

Recent developments in the generation of number states, $|n\rangle$, suggest their use in future communications systems. Since number states are orthogonal, in principle they can be detected without error. Furthermore, since number states represent energy, classical energy detectors, such as photon counters, provide optimum quantum measurements. Therefore, once the number states are generated and transmitted, their detection via classical photon-counting detectors is certain provided the channel does not introduce any appreciable losses. If the channel is lossy, then the number-state channel transforms into a classical erasure channel, where the only source of error is loss of the transmitted state.

For the case of number-state communications, capacity can be determined using a two-step process, where the external noise is first reduced to zero, resulting in a classical erasure channel. Then the channel losses are further reduced to zero, yielding the Shannon capacity.

In this section we consider several physical constraints and show their impact on the achievable throughput. We also show methods of constructing modulation codes to implement the constraint.

4.4.1 Shannon Capacity

Constraints on the allowable sequences of 0's and 1's may be described in a concise manner by a labelled graph. Figure 4-11 illustrates a graph describing the deadtime constraint in which each 1 is separated by at least one 0.

A graph $\mathcal{G}$ consists of a set of states $\mathcal{V}$ and a set of labeled edges $\mathcal{E}$. Each edge e is directed, with initial state $i(e) \in \mathcal{V}$ and terminal state $t(e) \in \mathcal{V}$. Sequences are generated by traversing edges in the graph and reading off, or concatenating, their labels to form words or finite strings of bits. The collection of all sequences presented by a graph is referred to as a *constrained system* L^S. We assume throughout that the graphs are *irreducible*, meaning that there is a path between any two states in the graph; *right-resolving*, meaning that the labels of paths beginning in the same state generate distinct words for all sufficiently

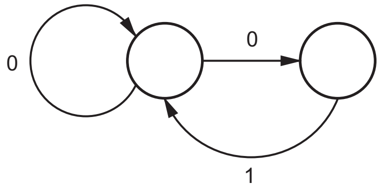

Fig. 4-11. A one-slot deadtime constraint graph.

long paths; and *primitive*, meaning that for some m, there is a path between any two states of length m. Most constraints of practical interest have an irreducible, right-resolving, primitive presentation.

A constrained code is an invertible mapping of unconstrained binary sequences into constrained binary sequences. The maximum rate of an invertible code that maps source sequences into sequences from the constrained system is referred to as the Shannon capacity of the system. It is the noiseless capacity of a channel defined by the graph in the sense that sequences are transmitted noiselessly, but are constrained to be those generated by paths on the graph. Letting $N(n, S)$ denote the number of distinct words of length n occurring in a constrained system S, the Shannon capacity is given by

$$C(S) \overset{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n} \log_2 N(n, S) \text{ bits/slot} \tag{4.4-1}$$

For example, the Shannon capacity of the system illustrated in Fig. 4-11 is 0.6942 bits/slot, which tells us that, if pulses must be separated by at least one non-pulsed slot, we are strictly limited to data rates less than 0.6942 bits/slot. There are several approaches to determining the capacity. We will present them here without proof.

4.4.1.1 Characterizing Capacity: Fixed Duration Edges. Suppose we describe the constraint with a graph wherein all edges correspond to a duration of one slot. The *adjacency matrix* of $\mathcal{G}$ is the matrix $A_{\mathcal{G}}$ with the u, vth entry containing the number of edges from state u to state v. For example $A_{\mathcal{G}}$ corresponding to Fig. 4-11 is

$$A_{\mathcal{G}} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \tag{4.4-2}$$

Let S be the system presented by $\mathcal{G}$. One can show

$$C(S) = \log_2 \big(\rho(A_{\mathcal{G}})\big) \text{ bits/slot} \tag{4.4-3}$$

where $\rho(A_{\mathcal{G}})$ is the spectral radius of $A_{\mathcal{G}}$, the largest magnitude eigenvalue of the adjacency matrix. For any non-negative matrix such as an adjacency matrix, the spectral radius is non-negative. For $\mathcal{G}$ that is irreducible and right-resolving, the spectral radius has a multiplicity of one. The spectral radius of Eq. (4.4-2) is $(1+\sqrt{5})/2$, the golden mean. Hence the capacity is 0.6942 bits/slot.

To determine the spectral radius for matrices of small dimension, we may find the eigenvalues and choose the largest. However, this is computationally prohibitive for matrices of large dimensions. In this case, one can find a good approximation to the spectral radius via the power method, e.g., [56, Exercise 8.5.16].

4.4.1.2 Characterizing Capacity: Variable Duration Edges. The previous characterization required each edge in the graph to correspond to one slot, or a constant duration. We may describe the constraint more concisely, and allow a broader class of constraints, by allowing edges to have variable durations. For example, Fig. 4-12 is an equivalent presentation of the constraint presented in Fig. 4-11. Let $\tau(e)$ be the duration of edge e, measured in slots. The *partition matrix* of $\mathcal{G}$ is the matrix with i, jth entry

$$B_{i,j}(s) = \sum_{e \in \mathcal{E}, i(e)=i, t(e)=j} e^{-s\tau(e)}$$

For example, the partition matrix corresponding to Fig. 4-12 is $B(s) = e^{-s} + e^{-2s}$. It can be shown that [57]

$$C(S) = \frac{s_0}{\ln(2)} \text{ bits/slot} \tag{4.4-4}$$

where s_0 is the solution to $\rho(B(s)) = 1$. The spectral radius $\rho(B(s))$ is strictly decreasing in s; hence, the solution is unique. For $B(s) = e^{-s} + e^{-2s}$, $s_0 = \ln((1+\sqrt{5})/2)$. Hence, the capacity of Fig. 4-12 is, again, 0.6942 bits/slot. Note that if all edges have duration 1, then $B(s) = \exp(-s)A_{\mathcal{G}}$, $s_0 = \ln(\rho(A_{\mathcal{G}}))$, and $C(S) = \log_2 \rho(A_{\mathcal{G}})$, in agreement with Eq. (4.4-3).

4.4.1.3 Characterizing Capacity: Probabilistic Characterization. We can consider the graph as describing a Markov source with each edge assigned a *transition probability* $p(e)$ such that

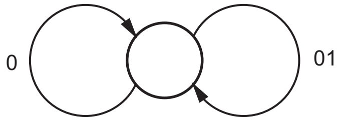

Fig. 4-12. A one-slot deadtime constraint graph, variable duration edges.

$$\sum_{e,i(e)=v} p(e) = 1 \ \forall v \in \mathcal{V}$$

Let $P_{u,v} = \sum_{e,i(e)=u,t(e)=v} p(e)$ be the *transition probability matrix* with the u, vth entry denoting the probability of a transition from u to v, given the source was in state u. For an irreducible graph, the $p(e)$'s define a unique stationary distribution μ on the states, the left eigenvector of the transition probability matrix

$$\mu^T = \mu^T P$$

normalized such that $\sum_{v \in \mathcal{V}} \mu(v) = 1$.

The *absolute transition probabilities* $q(e) = \mu(i(e))p(e)$ denote the probability that edge e is selected at an arbitrary time and satisfy $\sum_{e \in \mathcal{E}} q(e) = 1$. The information conveyed by transmitting the label of e is $-\log_2 q(e)$ bits, and the time required to send this information is $\tau(e)$. The average information that may be conveyed by this source per edge is the entropy rate of the source

$$H(\mathbf{q}) = -\sum_{e \in \mathcal{E}} q(e) \log_2 p(e) \text{ bits/edge}$$

However, edges have variable durations. The average duration of an edge is

$$T(\mathbf{q}) = \sum_{e \in \mathcal{E}} q(e)\tau(e) \tag{4.4-5}$$

hence, the information rate per slot is

$$\frac{H(\mathbf{q})}{T(\mathbf{q})} \text{ bits/slot}$$

Let $\mathbf{q}^*$ be the distribution that maximizes the information rate in bits per slot. Shannon [6] proved that this probabilistic notion of capacity is equivalent to the combinatorial notion given in Eq. (4.4-1), i.e.,

$$C(S) = \frac{H(\mathbf{q}^*)}{T(\mathbf{q}^*)} \tag{4.4-6}$$

It is informative to determine the maximizing distribution $\mathbf{q}^*$. Let ξ and η be the unique left and right eigenvectors of $B(s_0)$ associated with eigenvalue 1,

$$\xi B(s_0) = \xi$$

$$B(s_0)\eta = \eta$$

normalized such that $\xi\eta = 1$. The maximizing distribution is given by

$$p^*(e) = \frac{\eta_{t(e)}}{\eta_{i(e)}} e^{-s_0\tau(e)}$$

$$q^*(e) = \xi_{i(e)}\eta_{t(e)} e^{-s_0\tau(e)}$$

$$\mu^*(u) = \eta_u \xi_u$$

4.4.1.4 Characterizing Capacity: Energy Efficiency. In the prior sections, we defined capacity as the least upper bound on the efficiency of the modulation in bits per slot. We may also define capacity as the least upper bound on the efficiency of the modulation in bits per photon. The characterization from Sections 4.4.1.2 and 4.4.1.2 extend to this case in a straightforward manner by assigning edges a cost $\tau(e)$ in the appropriate units.

Redefine $\tau(e)$ to be the number of photons transmitted when edge e is traversed. Then the average number of photons per edge is given by Eq. (4.4-5), and the capacity in bits per photon is given by Eq. (4.4-6). For example, in Fig. 4-12, let the edge labeled ‘01’ cost K_s photons, and the edge labeled ‘0’ cost no photons. Then $B(s) = 1 + e^{-s}$, and s_0 is the solution to $e^{-s_0} = 0$. In this case, we can transmit at arbitrarily large bits/photons. One can see this is accomplished by letting the probability of using edge ‘01’ go to zero.

4.4.2 Constraints

We consider several constraints on laser transmission. Certain lasers require a recovery time after transmission of a pulse, imposing a minimum delay, or deadtime, between pulses. Timing recovery will impose a constraint on the maximum time that may elapse between pulses. Finally, any deep-space optical system will be subject to an average power constraint, or, equivalently for equal-power pulses, a duty-cycle constraint.

4.4.2.1 Dead Time. We first consider the general form of the deadtime constraint, presented in Fig. 4-13. The capacity, from Eq. (4.4-4) is

$$C(d) = \log_2(\lambda)$$

where λ is the largest positive root of

$$\lambda^{-(d+1)} + \lambda^{-1} - 1 = 0 \tag{4.4-7}$$

For small d, exact solutions may be found efficiently for Eq. (4.4-7). For large d, substitute $\lambda = e^{C(d)}$ and use the approximation $e^{-C(d)} \approx 1 - C(d)$, which yields $C(d)e^{(d+1)C(d)} \approx 1$ and thus

$$d + 1 \approx (d+1)C(d)e^{(d+1)C(d)}$$

or

$$C(d) \approx \frac{W(d+1)}{(d+1)\ln 2} \text{ bits/slot} \tag{4.4-8}$$

where $W(z)$ is the *productlog* function, which gives the solution for w in $z = we^w$.

Denote the two edges in the graph e_0 and e_{10^d}. Each transmission of e_0 can be thought of as an unpulsed slot in excess of the minimum required. For large d, the optimal probability of this transmission is

$$\begin{aligned} p(e_0) &\approx e^{W(d+1)/(d+1)} \\ &= \exp\big(-\exp\big(-W(d+1)\big)\big) \end{aligned}$$

The productlog function grows roughly like the log, so, to a rough approximation, the probability of inserting excess zeros grows as $e^{-1/(d+1)}$.

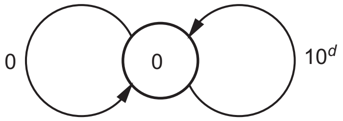

Fig. 4-13. Deadtime constraint.

4.4.2.2 Runlength. Any practical system will also limit the duration between pulses to aid in timing recovery. Let k denote the maximum number of consecutive slots without a pulse, or the *runlength* constraint. Figure 4-14 illustrates the deadtime and runlength constraint. The capacity, from Eq. (4.4-4), is

$$C(d,k) = \log_2(\lambda) \text{ bits/slot}$$

where λ is the largest positive root of

$$\lambda^{-(d+1)} + \lambda^{-(d+2)} + \cdots + \lambda^{-(k+1)} - 1 = 0$$

The subject of deadtime and runlength constraints has been extensively studied in the literature pertaining to magnetic and optical recording, e.g., [58], where modulation codes of this type are extensively used.

The energy efficient case has a simple solution. Let each transmission of a 1 cost a single photon, and let $C_E(d,k)$ be the energy-efficient deadtime and runlength capacity. We have

$$C_E(d,k) = \log_2 k - d + 1 \text{ bits/photon}$$

achieved by choosing each of the edges with equally likely probability.

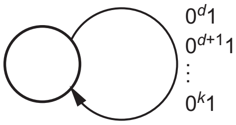

Fig. 4-14. Deadtime and runlength constraint.

4.4.3 Modulation Codes

A modulation code is an invertible mapping from unconstrained binary sequences into a system S. Let $R_{\mathcal{C}}(S)$ denote the rate of a modulation code $\mathcal{C}$ into the system S. Then $E_{\mathcal{C}(d)} \overset{\text{def}}{=} R_{\mathcal{C}}(S)/C(S)$ is the relative *efficiency* of the code, measuring how close the code rate is to the limit. Note that the efficiency of a code is measured relative to the constraint it is designed to satisfy.

There are well-known techniques to construct codes into a modulation system at rates arbitrarily close to capacity, e.g., [59,60]. However, for our parameter range, a straight-forward application of these approaches may be prohibitively complex. In the following sections we present some approaches that trade efficiency for complexity.

4.4.3.1 *M*-ary PPM with Deadtime. PPM may be made to satisfy a deadtime constraint by appending d non-pulsed slots after each PPM frame as illustrated in Fig. 4-15. A graph describing the allowable sequences for deadtime-PPM is illustrated in Fig. 4-16.

The rate is

$$R_{\text{ppm}}(d, M) = \frac{\log_2(M)}{M + d} \text{ bits/slot}$$

We will allow non-integer M in analysis to simplify expressions, since rounding has a negligible effect on rate for large d. For a given value of d we find M^*, the argument that maximizes $R_{\text{ppm}}(d, M)$, by solving $\partial R_{\text{ppm}}(d, M)/\partial M = 0$, which yields

$$\frac{\ln(M^*)}{M^* + d} = \frac{1}{M^*}$$

or

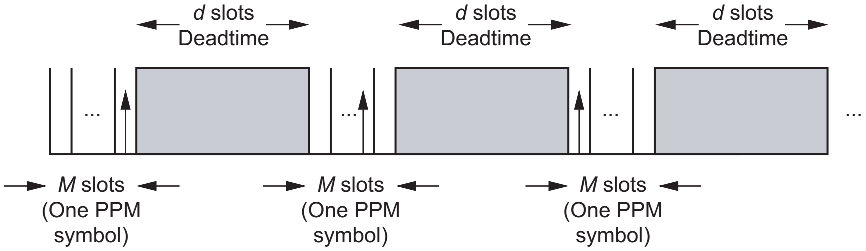
d slots
Deadtime
d slots
Deadtime
d slots
Deadtime
...
...
...
...
M slots
(One PPM
symbol)
M slots
(One PPM
symbol)
M slots
(One PPM
symbol)

Fig. 4-15. PPM signaling with a deadtime constraint.

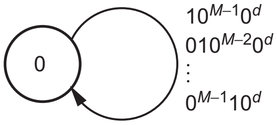
0
$10^{M-1}0^d$
$010^{M-2}0^d$
⋮
$0^{M-1}10^d$

Fig. 4-16. Deadtime-PPM.

$$M^* = \frac{d}{W(d/e)} \tag{4.4-9}$$

Noting that $R_{\text{ppm}}(d, M^*) = 1/(M^* \ln 2)$, the maximum rate is

$$R_{\text{ppm}}(d) = \frac{W(d/e)}{d \ln 2} \text{ bits/s}$$

By an application of L'Hôpital's rule, one can show $R_{\text{ppm}}(d)/C(d) \to 1$ as $d \to \infty$, i.e., deadtime-PPM achieves deadtime capacity in the limit of large d. However, for small to moderate d, significant gains in throughput over PPM are available.

4.4.3.2 *M*-ary DPPM with Deadtime. With deadtime-PPM, there are d unused slot positions in the transmitted signal following each frame—unused in the sense that they neither convey information nor are always necessary for satisfying the deadtime constraint. It would be more efficient to map $\log_2 M$ bits to a pulsed slot position and follow each pulse by exactly d non-pulsed slots as illustrated in Fig. 4-17. This signaling scheme is referred to as deadtime-differential-PPM (deadtime-DPPM), presented by Fig. 4-18.

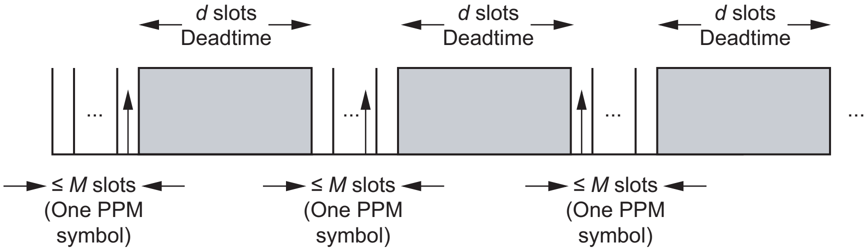
d slots Deadtime
d slots Deadtime
d slots Deadtime
...
...
...
...
≤ M slots (One PPM symbol)
≤ M slots (One PPM symbol)
≤ M slots (One PPM symbol)

Fig. 4-17. In DPPM signaling, the designated deadtime begins immediately after the pulse of the PPM symbol.

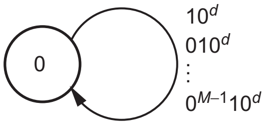
0
10^d
010^d
⋮
$0^{M-1}10^d$

Fig. 4-18. Deadtime-DPPM.

Since the duration of a codeword mapping to $\log_2 M$ bits is variable, deadtime-DPPM has an *average* rate

$$R_{\text{dppm}}(d, M) = \frac{\log_2(M)}{\dfrac{M+1}{2} + d} \text{ bits/slot}$$

For a given value of d we find M^*, the argument that maximizes $R_{\text{dppm}}(d, M)$, by solving $\partial R_{\text{dppm}}(d, M)/\partial M = 0$, which yields

$$M^* = \frac{2d+1}{W\left(\dfrac{2d+1}{e}\right)}$$

and maximum achievable rate

$$R_{\text{dppm}}(d) = \frac{2W\left(\dfrac{2d+1}{e}\right)}{(2d+1)\ln 2} \text{ bits/slot}$$

Since $R_{\text{dppm}}(d)$ is bounded above by $C(d)$ and below by $R_{\text{ppm}}(d)$, DPPM also achieves capacity in the limit of large d. DPPM is a low-complexity scheme that demonstrates significant throughput gains over PPM. However, there are practical issues with implementing variable-rate decoders. Hence we investigate an intermediate solution.

4.4.3.3 Synchronous Variable-Length Codes. The encoders considered so far have been either fixed rate or variable rate. Allowing a variable rate adds a degree of freedom in design, resulting in higher efficiency and/or lower complexity encoders. However, variable rate encoding and decoding has practical drawbacks. A compromise is to allow a synchronous rate, namely mappings of mp bits to mq bits, where p, q are fixed positive integers, and m is a positive integer that can vary. Methods of constructing synchronous variable-length codes were initially described in [61], and reviews of various approaches may be found in [58,60].

We describe a systematic procedure to construct synchronous encoders and decoders for (d, ∞) constraints. The procedure may be interpreted as a practical method of approaching rates of deadtime-DPPM.

Choose a rate $p/q < C(d)$ bits/slot. We desire a set of variable-length codewords $\mathcal{C} = \{c_1, c_2, \cdots, c_N\}$ such that any sequence formed by freely concatenating the codewords satisfies the constraint, the codeword lengths $l(c_i)$ are multiples of q, no codeword is the prefix of another (sufficient but not necessary to guarantee decodability), and the collection satisfies the Kraft (In)equality:

$$\sum_{c_i \in \mathcal{C}} 2^{-l(c_i)p/q} = 1 \tag{4.4-10}$$

We can use such a set to construct a synchronous variable-length code mapping unconstrained binary sequences into the constraint.

We detail one method to construct such a set that leads to a low-complexity encoder and decoder. The codewords are constructed as nodes on a binary tree. The root of the tree is the pattern 0^d. Branches with a label 1 are extended with zeros to the first length that is a multiple of q. At this point, the branch label is taken as a codeword. The tree is expanded until we have a set of codewords that satisfies Eq. (4.4-10). Figure 4-19 illustrates the procedure for $(d, k) = (16, \infty)$, $q = 7$.

The all-zeros pattern is not allowed as a codeword since allowing it reduces the minimum Euclidean distance from 2 to 1, the small gain in throughput does not offset the loss due to the smaller distance (allowing the all zeros codeword does yield significant throughput gains for small d), and a finite k constraint is desired for synchronization. The encoding and decoding may be done at a fixed rate by using encoders and decoders with appropriate memory. A simple encoder implementation exists if we allow variable-out-degree states.

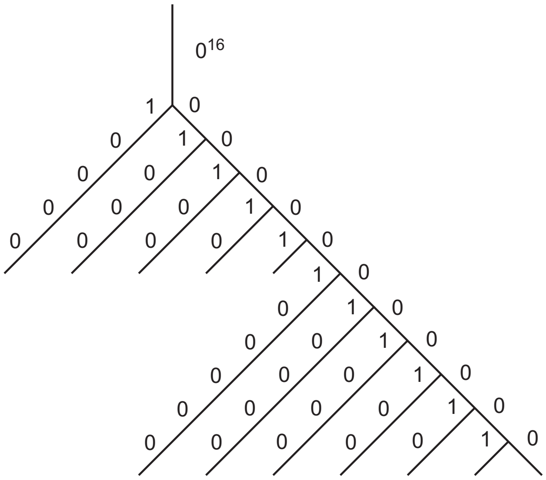

Fig. 4-19. Synchronous variable-length construction for $(d,k) = (16,\infty)$, $q = 7$.

4.5 Performance of Uncoded Optical Modulations

We now analyze the performance of the modulation schemes discussed in Section 4.3 when used on channels governed by the models described in Section 4.2. As in Fig. 4-2, let X denote the message to be sent on the channel and let Y denote the output of the detector. If we observe $Y = y$, the maximum-likelihood (ML) detector[1] produces the output

$$\hat{X} = \underset{x}{\operatorname{argmax}}\, f_{Y|X}(y|x) \tag{4.5-1}$$

where $f_{Y|X}(y|x)$ is the conditional probability density function or probability mass function of Y, given $X = x$. If X is a priori uniformly distributed, as it is for most cases of interest, the ML detector minimizes the probability of detector error $\Pr(\hat{X} \neq X)$. We shall assume a uniform a priori distribution throughout the rest of the chapter.

When X can take on only the values 0 or 1, it is often convenient to write the ML detector in terms of the log-likelihood ratio $\Lambda(y) = \ln L(y) = \ln[f_{Y|X}(y|1)/f_{Y|X}(y|0)]$. The log-likelihood ratio test is an equivalent way to write Eq. (4.5-1) [62]:

$$\hat{X} = \begin{cases} 1 & \text{if } \Lambda(y) > 0 \\ 0 & \text{if } \Lambda(y) \leq 0 \end{cases} \tag{4.5-2}$$

When $\Lambda(y) = 0$, Eq. (4.5-2) defines $\hat{X} = 0$, but the error rate would not be affected by defining $\hat{X}$ to be randomly selected from 0 and 1. As we shall see in the following sections, the log-likelihood ratio test often reduces to a simple threshold test of the form

$$\hat{X} = \begin{cases} 1 & \text{if } y > \tau \\ 0 & \text{if } y \leq \tau \end{cases} \tag{4.5-3}$$

---

[1] The use of "detector" here does not refer to just the physical recording of photons, but rather the more general algorithmic problem of identifying a transmitted signal given one or more observables.

4.5.1 Direct Detection of OOK on the Poisson Channel

The Poisson channel with $f_{Y|X}(k|0) = K_b^k e^{-K_b}/k!$ and $f_{Y|X}(k|1) = (K_s + K_b)^k e^{-(K_s+K_b)}/k!$ results in a log-likelihood ratio of

$$\Lambda(k) = \ln\left(\frac{(K_s + K_b)^k e^{-(K_s+K_b)}}{k!} \cdot \frac{k!}{K_b^k e^{-K_b}}\right) = k \ln\left(1 + \frac{K_s}{K_b}\right) - K_s \tag{4.5-4}$$

from which we can see that the ML rule becomes a threshold test as in Eq. (4.5-3), with threshold

$$\tau = \frac{K_s}{\ln\left(1 + \dfrac{K_s}{K_b}\right)} \tag{4.5-5}$$

Denoting the expression "greatest integer less than or equal to x" by $\lfloor x \rfloor$, the bit error rate (BER) can be expressed as

$$P_b = \frac{1}{2} - \frac{1}{2}\sum_{k=0}^{\lfloor \tau \rfloor}\big(f(k|0) - f(k|1)\big) \tag{4.5-6}$$

which is a finite sum of easily computed terms. When no background is present, then $K_b = 0$, the threshold in Eq. (4.5-5) becomes $\tau = 0$, and the summation reduces to a single term:

$$P_b = \frac{1}{2} e^{-K_s}$$

The performance of OOK on a Poisson channel is shown in Fig. 4-20 for various values of K_b. Discontinuities in the slope of the performance curves for $K_b > 0$ occur when the optimum threshold changes by an integer amount. A more general derivation of the optimum threshold for the case of unequal a priori probabilities can be found in [63].

The BER may be expressed in terms of the photon efficiency, as this provides a direct way to determine system efficiency when operating at a prescribed error probability [64]. Since each OOK symbol corresponds to one bit and the average photons per symbol is $K_s/2$, the photon efficiency of uncoded OOK signaling can be expressed as

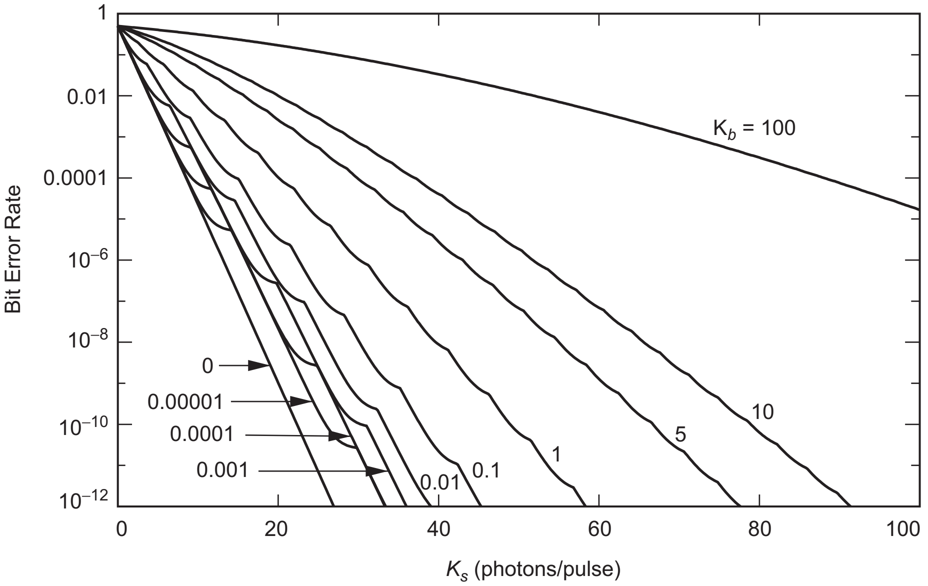

Fig. 4-20. BER versus signal level for uncoded OOK signaling on a Poisson channel, for various background levels.

$$\rho = \frac{2}{K_s} \text{ bits/photon} \tag{4.5-7}$$

This leads to the following alternate expression for BER, in terms of photon efficiency, when $K_b = 0$:

$$P_b = \frac{1}{2} e^{-2/\rho} \tag{4.5-8}$$

The result shows that operating at high values of ρ tends to result in high BERs. By generating plots of the error probability as a function of ρ, the photon efficiency of the system can be determined at any desired uncoded BER.

4.5.2 Direct Detection of PPM

At the receiver, the detected photons in each slot are counted. If hard decisions are used, the slot with the greatest photon count is declared to be the signal slot. This has been shown to be the ML decision for the channel models described in Section 4.2 [65]. On a continuous-output channel, the PPM symbol error probability P_s is the well-known performance [47,66] of an ML detector for M-ary orthogonal signaling:

$$P_s = 1 - \Pr\big(Y_1 = \max\{Y_1, \cdots, Y_M\} | X_1 = 1\big)$$

$$= 1 - \int_{-\infty}^{\infty} f_{Y|X}(y|1) \left[\int_{-\infty}^{y} f_{Y|X}(y'|0)dy'\right]^{M-1} dy \qquad (4.5\text{-}9)$$

In Eq. (4.5-9), observations in the $M - 1$ nonsignal slots are assumed to be independent. Equation (4.5-9) may be evaluated numerically by first producing a table lookup for the bracketed term, and then computing the outer integral numerically in the usual way.

When the detector outputs take on discrete values, there is a possibility of a tie for the maximum count. The solution for this problem has been derived in [7]. Suppose k photons are detected in the slot containing the pulse, l nonsignal slots also have count k, and the remaining nonsignal slots have count strictly less than k. Then the correct decision is made with probability $1/(l+1)$. Otherwise, an error is made. By summing over all possible values of k and l, it follows that

$$P_s = 1 - \sum_{k=0}^{\infty} \sum_{l=0}^{M-1} \Pr \begin{bmatrix} \text{correct decision when} \\ l \text{ nonsignal slots tie} \\ \text{the signal slot for the} \\ \text{maximum count} \end{bmatrix} \times \Pr \begin{bmatrix} \text{exactly } l \text{ of } M-1 \\ \text{nonsignal slots have} \\ \text{value } k, \text{ all others} \\ \text{smaller} \end{bmatrix} \qquad (4.5\text{-}10)$$

$$\times \Pr \begin{bmatrix} \text{signal slot} \\ \text{has value } k \end{bmatrix} \qquad (4.5\text{-}11)$$

$$= 1 - \sum_{k=0}^{\infty} \sum_{l=0}^{M-1} \frac{1}{l+1} \binom{M-1}{l} f_{Y|X}(k|0)^l F_{Y|X}(k-1|0)^{M-l-1} f_{Y|X}(k|1) \qquad (4.5\text{-}12)$$

where $f_{Y|X}(k|1)$ and $f_{Y|X}(k|0)$ denote the conditional probabilities that a received slot $Y_i = k$ when a 1 (pulse) or 0 (no pulse), respectively, is transmitted in the slot, and $F_{Y|X}(k|1)$ ($F_{Y|X}(k|0)$) denote the cumulative distributions. An extension of Eq. (4.5-12) to n-pulse PPM, $n \geq 2$, is straightforward, and involves a triple summation in place of the double summation. After some algebraic manipulation, this can be rewritten in a single summation as [67]

$$P_s = 1 - \frac{1}{M}\sum_{k=0}^{\infty} L(k)\left(F_{Y|X}(k|0)^M - F_{Y|X}(k-1|0)^M\right) \tag{4.5-13}$$

where $L(k)$ is the likelihood ratio. On a noiseless channel, $f_{Y|X}(0|0) = 1$, and the erasure probability is $f_{Y|X}(0|1)$, and thus, Eq. (4.5-13) can be simplified to

$$P_s = \frac{(M-1)f_{Y|X}(0|1)}{M} \tag{4.5-14}$$

Once the PPM symbol is detected, it is mapped to a string of $\log_2(M)$ bits via the inverse of the encoding mapping (see Fig. 4-6). There are $M/2$ symbol errors that will produce an error in a given bit in the string, and there are $M-1$ unique symbol errors. Thus, assuming all symbol errors are equally likely, the resulting bit error rate is

$$P_b = \frac{M}{2(M-1)}P_s \tag{4.5-15}$$

where P_s is given by Eq. (4.5-9) or Eq. (4.5-13).

4.5.2.1 Poisson Channel. For $K_b > 0$, the SER in Eq. (4.5-13) becomes

$$P_s = 1 - \sum_{k=0}^{\infty}\left(1 + \frac{K_s}{K_b}\right)^k \frac{e^{-K_s}}{M}\left(F_{Y|X}(k|0)^M - F_{Y|X}(k-1|0)^M\right) \tag{4.5-16}$$

where $F_{Y|X}(k|0) = \sum_{m=0}^{k} K_b^k e^{-K_b}/k!$. When $K_b = 0$, from Eq. (4.5-14) we have

$$P_s = \frac{(M-1)e^{-K_s}}{M} \tag{4.5-17}$$

and from Eq. (4.5-15) we see that $P_b = (1/2)e^{-K_s}$, which is independent of M and equal to that of OOK. This is shown in Fig. 4-21(a). As K_b increases, the dependence on M grows and performance for each M degrades, as seen in Figs. 4-21(b) and 4-21(c). It is not appropriate to interpret performance versus K_s as a measure of power efficiency, however. The average transmitter power is proportional to K_s/M photons per slot. Whereas low values of M in Figs. 4-21(a) through 4-21(c) produce a lower BER compared to high values of M, the situation is reversed in Figs. 4-21(d) through 4-21(f). The performance is also shown for bits/photon in Figs. 4-21(g) through 4-21(i), where it is also seen that

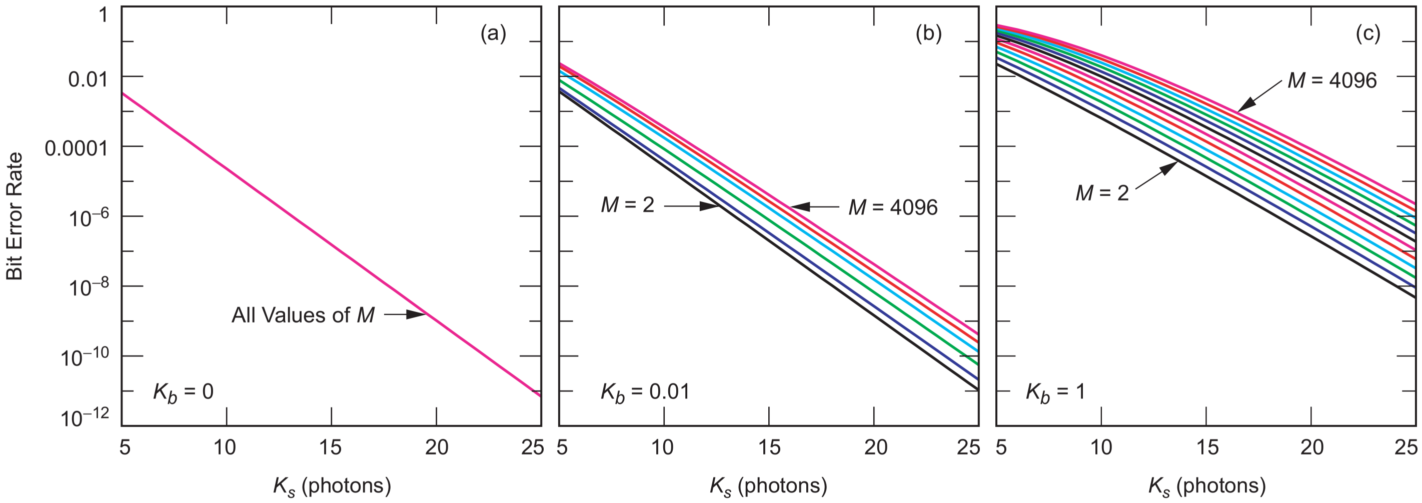

Fig. 4-21. BER of uncoded *M*-PPM signaling on a Poisson channel, versus (a-c) K_s photons/PPM symbol, (d-f) $P_{av} = K_s/M$ photons/slot, and (g-i) $\log_2 M/K_s$ bits/photon, for $M = 2$ to 4096 and $K_b = 0, 0.01, 1$.

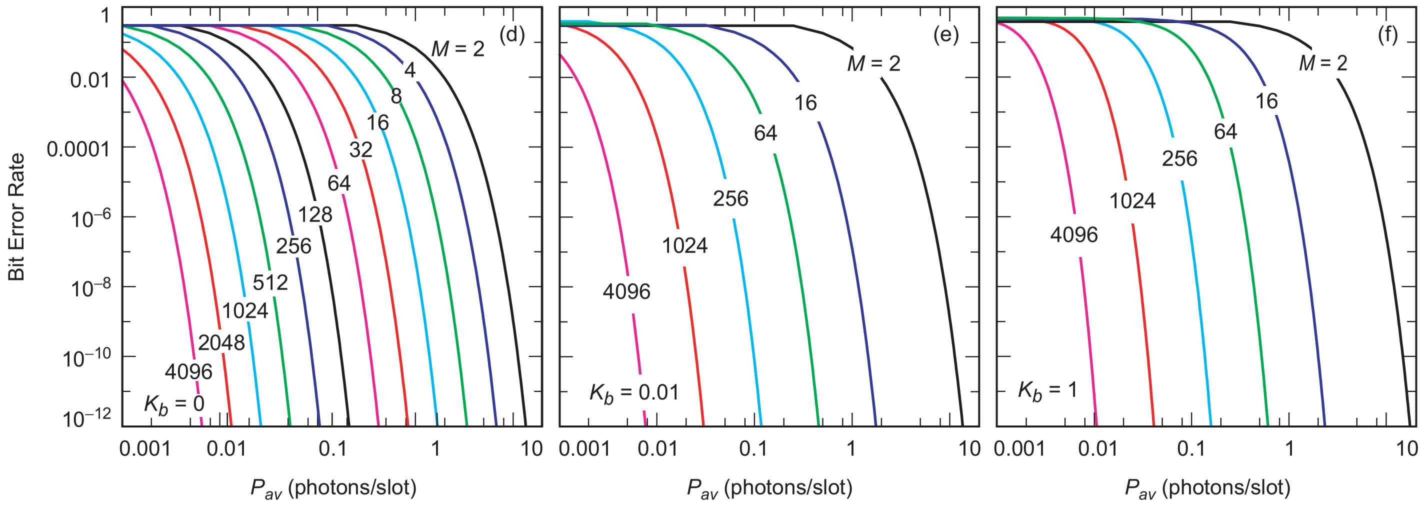

Fig. 4-21 (cont'd.).

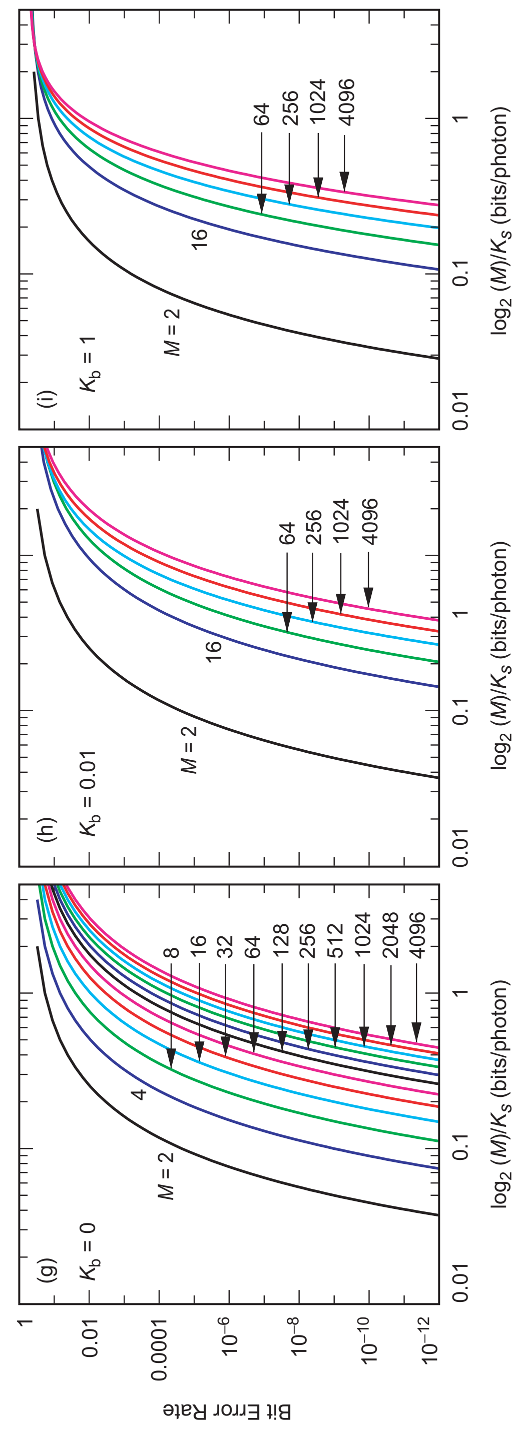

Fig. 4-21 (cont'd.).

high values of M are more photon efficient. Note that there is an average power difference of about 30 dB between $M = 2$ and $M = 4096$ [Figs. 4-21(d) through 4-21(f)], and a photon efficiency difference of about 10 dB [Figs. 4-21(g) through 4-21(i)]. The performance is also shown in terms of the slot SNR $\beta = K_s^2/K_b$ in Figs. 4-22(a) and 4-22(b), where we see little dependence on M; when plotted in terms of the bit SNR $\beta_b = K_s^2/(2K_b \log_2 M)$, the 30-dB gap manifests itself again, as seen in Figs. 4-22(c) and 4-22(d).

Expressing the BER in terms of the photon efficiency can be accomplished analytically when $K_b = 0$, by writing K_s in terms of ρ and substituting into Eq. (4.5-17), as originally shown in [64]:

$$P_b = \frac{1}{2} e^{-(\log_2 M/\rho)} = \frac{1}{2} M^{-(\log_2 e/\rho)} = \frac{1}{2} M^{-1.44/\rho} \tag{4.5-18}$$

4.5.2.2 AWGN Channel. The probability of symbol error is given by Eq. (4.5-9), which becomes

$$P_s = 1 - \int_{-\infty}^{\infty} \frac{1}{\sigma_1} \phi\left(\frac{x - m_1}{\sigma_1}\right) \Phi\left(\frac{x - m_0}{\sigma_0}\right)^{M-1} dx \tag{4.5-19}$$

$$= 1 - \int_{-\infty}^{\infty} \sqrt{\frac{\gamma}{\beta + \gamma}} \phi\left(\sqrt{\frac{\gamma}{\beta + \gamma}} \left(v - \sqrt{\beta}\right)\right) \Phi(v)^{M-1} dv \tag{4.5-20}$$

where $\phi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ is the standard normalized Gaussian probability density function, $\Phi(x)$ is its cumulative distribution function, and, as defined in Section 4.2.3, $\beta = (m_1 - m_0)^2/\sigma_0^2$ and $\gamma = (m_1 - m_0)^2/(\sigma_1^2 - \sigma_0^2)$. If $\sigma_1^2 = \sigma_0^2$, then $\gamma = \infty$ and Eq. (4.5-20) simplifies to

$$P_s = 1 - \int_{-\infty}^{\infty} \phi\left(x - \sqrt{\beta}\right) \Phi(x)^{M-1} dx \tag{4.5-21}$$

The BER P_b is then given by Eq. (4.5-15) and can be expressed in terms of the bit SNR $\beta_b = \beta/(2R_c)$, where in uncoded M-PPM there are $R_c = \log_2 M$ bits per PPM symbol.

For the case of homodyne detection, the BER can also be expressed in terms of ρ, as shown in [68]:

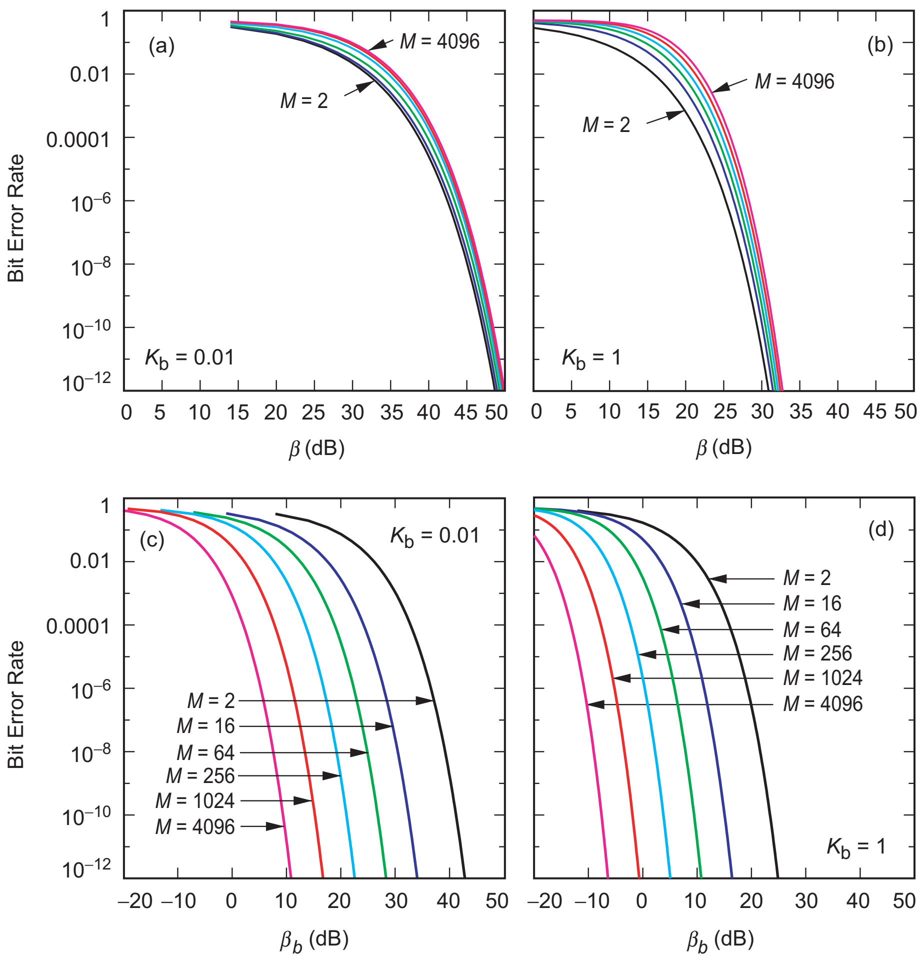

Fig. 4-22. BER of uncoded *M*-PPM signaling on a Poisson channel versus (a-b) slot SNR $\beta = K_s^2/K_b$ and (c-d) bit SNR $\beta_b = K_s^2/(2K_b \log_2 M)$, for $M = 2$ to 4096 and $K_b = 0.01, 1$.

$$P_s = 1 - \int_{-\infty}^{\infty} dx \frac{e^{-x^2}}{\sqrt{\pi}} \left[1 - Q\left(x\sqrt{2} + 2\sqrt{\frac{\log_2 M}{\rho}}\right)\right]^{M-1} \tag{4.5-22}$$

The limiting form of the bit error probability, as the number of symbols (hence the dimension of the symbol set) grows without bound, is in [68]. Using Eq. (4.5-15) to relate P_s to P_b, we get

$$\lim_{M\to\infty} P_b = \begin{cases} \frac{1}{2}; & \frac{2}{\rho} < \ln 2 \\ 0; & \frac{2}{\rho} > \ln 2 \end{cases} \tag{4.5-23}$$

It follows that arbitrarily low error probabilities can be achieved as M approaches infinity, as long as the inequality $\rho < 2/\ln 2 \approx 2.89$ bits/photon is satisfied.

4.5.3 Direct Detection of Combined PPM and WSK

Combined PPM and WSK is an orthogonal signaling scheme. Therefore, its performance is the same as PPM, but with NM dimensions instead of M. The probability of correct detection is given by Eqs. (4.5-9) and (4.5-13), but with M replaced by NM.

The information efficiency of this combined modulation scheme is $\rho = (\log_2 NM)/K_s = (\log_2 N + \log_2 M)/K_s$ bits/photon, but the photon rate remains $n_s = K_s/T$ photons/second. Hence the data rate is

$$R = \frac{\log_2 NM}{K_s}\frac{K_s}{T} = \frac{\log_2 N + \log_2 M}{T} \text{ bits/second} \tag{4.5-24}$$

For a given bandwidth, the throughput of combined PPM and WSK is higher than PPM alone. However, this modulation scheme requires N detectors, one for each wavelength, instead of just one.

4.5.4 Performance of Modulations Using Receivers Based on Quantum Detection Theory

We now present a description of receivers based on quantum detection theory, along with some specific examples of optical modulation formats that are of potential interest for deep-space applications.

4.5.4.1 Receivers Based on Quantum Detection Theory. In 1973, Kennedy described an ideal coherent receiver with significant communication advantages over ideal heterodyning, homodyning, or direct detection receivers [69]. Kennedy's receiver adds a local field prior to photodetection that exactly negates the received field corresponding to one of the possible messages. The photodetector is guaranteed (ideally) to record zero photons in the case of this message. For the other message(s), the photodetector responds to the coherent sum of the signal field and the local field. For the binary coherent state detection problem, Kennedy's receiver obtains a 3-dB improvement over ideal homodyne or direct detection and a 6-dB improvement over ideal heterodyne detection.

Dolinar proposed a generalization of Kennedy's receiver structure [70] and optimized it to obtain a realization of an exactly optimum quantum measurement for the case of binary coherent state signals [70]. The optimum receiver for the binary case eventually nulls the field corresponding to the message that is ultimately selected as more likely. Initially, however, the receiver does not preselect one of the two messages for nulling. The added local field changes abruptly with each observed count from the photodetector as the message likelihoods are updated.

Dolinar also described a near-optimum conditionally nulling receiver for the M-ary PPM detection problem [71]. This receiver sequentially, slot by slot, based on prior observed counts, decides either to add or not add a local field designed to null the signal field that corresponds to a pulse in the current slot.

Since an ideal nulling or conditionally nulling receiver is a mathematical artifact that can never be precisely realized in practice, non-ideal versions that permit phase errors in the nulling signal have also been modeled and analyzed by Dolinar [71] and Vilnrotter and Rodemich [63].

The performance of the optimum quantum receiver has been determined in the literature, particularly for the cases of noiseless reception of binary and higher-dimensional signals and of noisy reception of binary signals. The performance literature has recently extended to higher dimensional signals operating in the presence of noise. A summary of these results follows.

4.5.4.1.1 Quantum Detection with Binary Decisions. The problem of determining which of two possible quantum states is present, representing binary hypotheses, has been addressed in detail for the case of both noisy and noiseless reception [3,5,75]. Under hypothesis H_0, the received electromagnetic field is in a mixture of states governed by the density operator ρ_0, and under hypothesis H_1, it is governed by ρ_1. Assume equal a priori probabilities, so that $\Pr(H_0) = \Pr(H_1) = 1/2$. Suppose a receiver applies the "detection operators" defined in terms of an appropriate set of orthonormal basis states $\{|b_n\rangle\}$:

$$\begin{aligned}\Pi_0 &= \sum_{n:\, b_n \in I_0} |b_n\rangle\langle b_n| \\ \Pi_1 &= \sum_{n:\, b_n \in I_1} |b_n\rangle\langle b_n|\end{aligned} \tag{4.5-25}$$

These detection operators effectively partition the outcome space R into two disjoint regions I_0 and I_1, such that $R = I_0 \cup I_1$, $I_0 \cap I_1 = \emptyset$. The detection operators are projection operators that partition the outcome space into an exhaustive set of disjoint decision regions. For binary detection the outcome space contains only two regions, $\Pi_0 + \Pi_1 = \mathbf{1}$; hence only one of the detection operators needs to be applied. If the outcome of the measurement falls within the region I_1, H_1 is selected. Otherwise, the receiver chooses H_0.

The conditional probabilities of a correct decision are [3,5,75]

$$\begin{aligned}\Pr(C|H_1) &= Q_d = \mathrm{Tr}\,[\rho_1 \Pi_1] \\ \Pr(C|H_0) &= 1 - Q_0 = 1 - \mathrm{Tr}\,[\rho_0 \Pi_1]\end{aligned} \tag{4.5-26}$$

It follows that, for equally probable signals, the average probability of correct decision is

$$\begin{aligned}P_c &= \frac{1}{2}\left\{\mathrm{Tr}\,[\rho_1 \Pi_1] + 1 - \mathrm{Tr}\,[\rho_0 \Pi_1]\right\} \\ &= \frac{1}{2}\left\{1 + \mathrm{Tr}\,[(\rho_1 - \rho_0)\Pi_1]\right\}\end{aligned} \tag{4.5-27}$$

and the average probability of error can be expressed as

$$P_b = 1 - P_c = \frac{1}{2}\left\{1 - \mathrm{Tr}\big[(\rho_1 - \rho_0)\Pi_1\big]\right\} \tag{4.5-28}$$

It is clear from the form of Eq. (4.5-28) that the probability of error is minimized by choosing the projection operator Π_1 to maximize the quantity $\mathrm{Tr}\,[\rho_1 - \rho_0)\Pi_1]$.

4.5.4.1.2 The Optimum Measurement for Binary Detection. Helstrom demonstrated that the detection operator which maximizes the quantity $\mathrm{Tr}\,[(\rho_1 - \rho_0)\Pi_1]$ is of the form [5]

$$\Pi^* = \sum_{k:\eta_k \geq 0} |\eta_k\rangle\langle\eta_k| \tag{4.5-29}$$

where $|\eta_k\rangle$ are the eigenvectors and η_k the eigenvalues of the operator $(\rho_1 - \rho_0)$, that is,

$$(\rho_1 - \rho_0)|\eta_k\rangle = \eta_k|\eta_k\rangle \tag{4.5-30}$$

The probability of error for this optimum projector is

$$\begin{aligned} P_b^* &= \frac{1}{2}\left\{1 - \mathrm{Tr}\left[(\rho_1 - \rho_0)\Pi^*\right]\right\} \\ &= \frac{1}{2}\left\{1 - \sum_{k:\eta_k > 0} \eta_k\right\} \end{aligned} \tag{4.5-31}$$

as shown in [3,5,75]. For "pure states" represented by $|\psi_0\rangle$ and $|\psi_1\rangle$, the density operators are $\rho_0 = |\psi_0\rangle\langle\psi_0|$ and $\rho_1 = |\psi_1\rangle\langle\psi_1|$. The eigenvectors of the operator $(\rho_1 - \rho_0)$ are now linear combinations of $|\psi_0\rangle$ and $|\psi_1\rangle$, yielding an expression of the form

$$|\eta_k\rangle = z_{k0}|\psi_0\rangle + z_{k1}|\psi_1\rangle, \qquad k = 0,\ 1 \tag{4.5-32}$$

where the coefficients are determined by substituting into Eq. (4.5-30) and equating coefficients. The optimum projection operator is simply $\Pi^* = |\eta_1\rangle\langle\eta_1|$, and its application yields the detection and false-alarm probabilities

$$Q_d = \mathrm{Tr}\left[\rho_1\Pi^*\right] = \frac{1}{2}\left[1 + \sqrt{1 - |\langle\psi_1|\psi_0\rangle|^2}\right] \tag{4.5-33}$$

$$Q_0 = \mathrm{Tr}\left[\rho_0\Pi^*\right] = \frac{1}{2}\left[1 - \sqrt{1 - |\langle\psi_1|\psi_0\rangle|^2}\right] \tag{4.5-34}$$

The probability of correct detection becomes

$$P_c^* = \frac{1}{2}\left[1 + \sqrt{1 - |\langle\psi_1|\psi_0\rangle|^2}\right] \tag{4.5-35}$$

while the probability of error is

$$P_b^* = 1 - P_c^* = \frac{1}{2}\left[1 - \sqrt{1 - |\langle\psi_1|\psi_0\rangle|^2}\right] \tag{4.5-36}$$

Thus, for the case of binary decisions, the error probability associated with the optimum projection measurement depends on the overlap between the two quantum states.

4.5.4.1.3 Orthogonal Signal States. With orthogonal signal states, $\langle\psi_0|\psi_1\rangle = 0$. Substituting into Eq. (4.5-36), we find that $P_b^* = 0$. Therefore, truly error-free communication could be achieved if practical orthogonal signal states could be generated, for example, if the number states $|\psi_0\rangle = |0\rangle$ and $|\psi_1\rangle = |1\rangle$ could somehow be prepared [75].

4.5.4.2 Performance of Representative Modulations. Performance of the following detection schemes for binary modulation has been described in [75], repeated here for reference.

4.5.4.2.1 On–Off Keying. For on–off keying we let

$$|\psi_0\rangle = |0\rangle$$

and

$$|\psi_1\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{(n!)^{1/2}}|n\rangle$$

where the average number of photons is $K_s = |\alpha|^2$ for the optical pulse, and zero for the ground state. The squared magnitude of the overlap between the two states is $|\langle\psi_0|\psi_1\rangle|^2 = e^{-|\alpha|^2}$, yielding $P_b^* = 1/2\left[1 - \sqrt{1 - e^{-|\alpha|^2}}\right] = 1/2\left[1 - \sqrt{1 - e^{-K_s}}\right]$ for optimum quantum detection. In comparison, the error probability for photon-counting detection was shown to be $P_b = (1/2)e^{-|\alpha|^2} = (1/2)e^{-K_s}$.

4.5.4.2.2 Optical BPSK. In this case of optical BPSK, we have two coherent states with the same average photon energy, but π radians out of phase. The signal states are defined as $|\psi_0\rangle = |\alpha\rangle$ and $|\psi_1\rangle = |\beta\rangle$ where the complex amplitudes are related as $\beta = -\alpha$, and the average number of signal photons is $K_s = |\alpha|^2$. Therefore, $|\langle\psi_0|\psi_1\rangle|^2 = e^{-4K_s}$, and the error probability becomes $P_b^* = 1/2\left[1 - \sqrt{1 - e^{-4K_s}}\right]$. A physically realizable receiver structure that achieves this error probability has been devised and analyzed by Dolinar [70].

4.5.4.2.3 Near-Optimum Optical BPSK. Exponentially optimum (or "near optimum") performance can be obtained by adding a local field of the same amplitude, in phase, to the received field followed by photon counting, as shown by Kennedy [69]. With this technique, the field amplitude under one hypothesis is shifted to the ground state, but doubled under the other. The error probability for this near-optimum detection scheme is $P_b = (1/2)e^{-4|\alpha|^2} = (1/2)e^{-4K_s}$.

4.5.4.2.4 Coherent Detection: The Classical Limit. If we add a local field of great amplitude, in phase, to the received field and detect the resulting sum field using classical energy detection, then the performance of the classical coherent optical detector is obtained. For optical BPSK modulation, the error probability for this receiver is given by [72] $P_b = Q\left(\sqrt{4|\alpha|^2}\right) = Q\left(\sqrt{4K_s}\right)$, where $Q(x) \equiv (1/\sqrt{2\pi})\int_x^\infty e^{-y^2/2}dy$. Performance curves for the various modulation formats and detection options are shown in Fig. 4-23.

4.5.4.2.5 Optimum Binary Detection in the Presence of Background Radiation. When background radiation is present, the received field is in a mix-

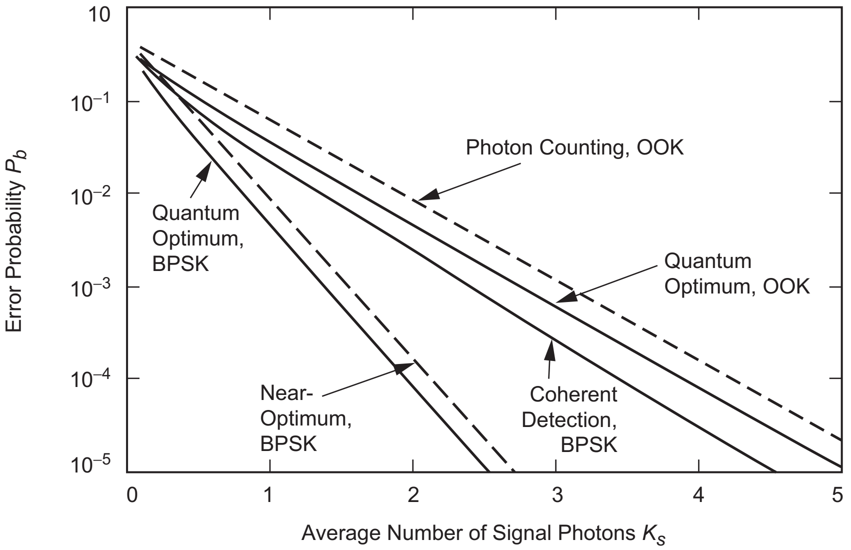

Fig. 4-23. Performance of binary optical receivers versus average signal photons per slot, assuming no background radiation.

ture of states described by an appropriate density operator. The probability of error can be calculated by finding the significant eigenvalues of the difference operator $(\rho_1 - \rho_0)$ and applying the formula for the error probability given by Eq. (4.5-31). Equivalently, the matrix representation of the difference operator can be diagonalized, the projection operator that selects only the positive eigenvalues applied, and the trace of the resulting diagonal matrix determined.

The following analysis parallels the derivation in [3] for on–off keying with noise. For this scenario, a density matrix ρ_1 can be defined for the signal and noise case. The controllable values are the size of the matrix, the average number of noise photons, and the average number of signal photons. The elements of this density matrix in the number state basis is, for $m \geq n$,

$$\langle n|\rho_1|m\rangle = (1-v)\sqrt{\frac{n!}{m!}}v^m \left(\frac{\mu^*}{N}\right)^{m-n} e^{\left(-(1-v)|\mu|^2\right)} L_n^{m-n}\left[-(1-v)\frac{|\mu|^2}{v}\right]$$

and for $m < n$,

$$\begin{aligned} \langle n\,|\rho_1|\,m\rangle &= \langle m\,|\rho_1|\,n\rangle^* \\ v &= \frac{N}{(N+1)} \end{aligned} \tag{4.5-37}$$

where N represents the average number of noise photons, μ is the complex envelope of the signal, $|\mu|^2$ is the average number of signal photons, and $L_n^{m-n}(x)$ is a Laguerre polynomial. A density operator ρ_0 for the null hypothesis, or noise-only case, can be generated using

$$\rho_0 = \sum_{n=0}^{\infty}(1-v)v^n\,|n\rangle\,\langle n| \tag{4.5-38}$$

Next, the resulting difference matrix is diagonalized using the formula $[\langle\eta_k|\rho_1 - \rho_0|\eta_k\rangle]$. The probability of error is found from this diagonalized matrix. The probability of correct detection is found by adding all the positive diagonal terms, yielding the probability of error

$$P_b = 1 - P_c = \frac{1}{2}\left\{1 - \sum_{k:\eta_k>0}\eta_k\right\} \tag{4.5-39}$$

For comparison, the performance of classical coherently detected BPSK signals observed in the presence of noise is called "threshold detection," with error probability $P_b = Q\left(\sqrt{4K_s/(2N+1)}\right)$, which is seen to be similar to the expression for the noiseless case, but with K_s replaced by $K_s/(2N+1)$. Note that for high background levels, the performance of the classical threshold detector is nearly as good as that of the optimum quantum detector. Hence, the physically realizable classical detector is a good approximation to the optimum quantum detector under conditions of high background radiation.

4.5.4.2.6 Multiple Hypotheses: Orthogonal Envelopes. The performance of the optimum quantum receiver for the case of equal-energy, equally probable signals is considered [3,5,75]. The signals are assumed to have orthogonal classical envelopes, which means that the classical complex envelopes $S_k(t)$ obey the condition

$$\int_0^T S_k^*(t) S_m(t)\,dt = 0, \quad k \neq m \tag{4.5-40}$$

In this case there are assumed to be M hypotheses, represented by M orthogonal classical envelopes modulating electromagnetic plane waves normally incident on the receiving aperture, and with temporal variation proportional to $S_k(t)$. PPM is an example of this modulation format where a single pulse is placed in one of M consecutive slots. The kth aperture-field mode is assumed to be matched to the kth signal, such that when the kth signal is present, the state of the aperture field is in a coherent state described by the "product state"

$$|\alpha_k\rangle = \prod_{j=1}^{M} |\alpha_{k,j}\rangle = |\alpha_{k,1}\rangle|\alpha_{k,2}\rangle \cdots |\alpha_{k,M}\rangle \tag{4.5-41}$$

where each of the $|\alpha_{k,j}\rangle$ is a coherent state associated with an individual mode, with $|\alpha_{k,j}| = |\alpha|\delta_{k,j}$, and where $K_s = |\alpha|^2$ is the average number of signal photons in each signal.

Since the M product states are linearly independent, the optimum strategy for minimizing the average probability of error is to project the received signal state onto M orthonormal measurement states spanning the same subspace and select the signal corresponding to the measurement state with the greatest projection. The state-space interpretation is similar to the binary case; because of symmetry, the M orthogonal measurement states are aligned with the M signal states in such a way as to maximize the projection of each state onto the corresponding measurement state, thus minimizing the probability of error. It has been shown [70] that for equally likely signals, the minimum symbol error probability for the optimum quantum receiver is

$$P_s = \frac{M-1}{M^2} = \left[\sqrt{1+(M-1)e^{-K_s}} - \sqrt{1-e^{-K_s}}\right]^2 \tag{4.5-42}$$

The error probability is given by $P_s = 1/2\left[1 - \sqrt{1 - |\langle\psi_1|\psi_2\rangle|^2}\right]$ as before, but now $|\langle\psi_1|\psi_2\rangle|^2 = e^{-2K_s}$, as compared to e^{-K_s} for the corresponding overlap with on–off keying, and e^{-4K_s} for BPSK signals. For comparison, the symbol error probability achieved by the photon-counting receiver in the absence of background radiation is given by Eq. (4.5-17). These results are summarized in [75].

For high-dimensional signaling and modest error probabilities on the order of $P_s \cong 10^{-3}$, conventional photon counting performs approximately 1.3 dB worse than the optimum quantum receiver in the absence of background radiation, but this performance gap increases at the lower error probabilities.

The performance of PPM and other high-dimensional signal envelopes in the presence of noise is not known in general; however, numerical techniques have been developed recently and applied to three-dimensional signals [73–75]. The performance of the optimum quantum receiver for the case of ternary signals in the presence of noise has been obtained. This signaling scheme is defined by three signal states, namely, $|\alpha\rangle$, $|-\alpha\rangle$, and $|0\rangle$. Further work is required to extend this approach to arbitrary dimensional signal sets operating in the presence of external noise.

4.6 Optical Channel Capacity

Shannon demonstrated in [76] that for any communications channel, as long as the rate of transmitting information is less than some constant, C, it is possible to make the average error probability arbitrarily small by coding messages into a large enough class of signals. The constant C is called the *capacity* of the channel. Characterizing the capacity of the optical channel provides a useful bound on the data rates achievable with any modulation and coding scheme, thus serving as a benchmark for assessing the performance of a particular design [10–20,22–27,43,48].

The capacity will be a function of the received optical signal and noise powers, the modulation, and the detection method. In this section, we assume the modulation is PPM. The loss in capacity by restricting the modulation to PPM is small (at most a few tenths of a dB) in the low average power regime where the deep-space optical channel currently operates.

We divide the capacity into two categories depending on the type of information provided to the decoder by the receiver. In one case, the receiver makes estimates of each PPM symbol, passing these estimates, or *hard* decisions, on to the decoder. In this case, the (hard-decision) capacity may be expressed as a function of the probability of symbol error, P_s, derived for several channel models in Section 4.5.

In the second case, the receiver makes no explicit symbol decision, but passes on slot counts (integrals of the received signal in each slot), or *soft* decisions, directly to the decoder. In this case, the (soft-decision) capacity may be expressed as a function of the channel statistic $f_{Y|X}$, presented for several channel models in Section 4.2.2. The soft-decision capacity is at least as large as the hard-decision capacity because the slot counts provide additional information to the decoder.

4.6.1 Capacity of the PPM Channel: General Formulas

The hard-decision PPM channel is an M-ary input, M-ary output, symmetric channel with capacity given by [66]

$$C = \log_2 M + (1 - P_s)\log_2(1 - P_s) + P_s \log_2\left(\frac{P_s}{M-1}\right) \text{ bits/PPM symbol} \tag{4.6-1}$$

where P_s is the probability of incorrect PPM symbol detection. The function P_s is provided for several channel models in Section 4.5.

The soft-decision capacity is given by

$$C = E_{\mathbf{Y}} \log_2\left[\frac{ML(Y_1)}{\sum_{j=1}^{M} L(Y_j)}\right] \text{ bits/PPM symbol} \tag{4.6-2}$$

an expectation over $\mathbf{Y}$, where $L(y) = f_{Y|X}(y|1)/f_{Y|X}(y|0)$ is the channel likelihood ratio, and the Y_j have density $f_{Y|X}(y|1)$ for $j = 1$ and density $f_{Y|X}(y|0)$ otherwise.

The M-fold integration in Eq. (4.6-2) is often intractable. However, it is straightforward to approximate the expectation by a sample mean. A quick approximation follows from the lower bound

$$C \geq E \log_2 \left[\frac{M}{1 + \dfrac{M-1}{L(Y_1)}} \right] \text{ bits/PPM symbol} \qquad (4.6\text{-}3)$$

which is a good approximation for large M, reducing the M-fold integration (or set of M-dimensional vector samples) needed to evaluate Eq. (4.6-2) to a one-dimensional integral (or set of scalar samples).

4.6.2 Capacity of Soft-Decision PPM: Specific Channel Models

4.6.2.1 Poisson Channel. We consider first the Poisson channel. The behavior of the case $K_b = 0$ has a particularly simple form. When $K_b = 0$, we have from Eqs. (4.2-17) and (4.2-18)

$$L(k) = \begin{cases} e^{-K_s} & k = 0 \\ \infty & k > 0 \end{cases}$$

and Eq. (4.6-2) reduces to

$$C = (\log_2 M)\left(1 - e^{-K_s}\right) \text{ bits/PPM symbol}$$

When $K_b = 0$, only signal photons are detected. If any signal photons are detected, the signal is known exactly. If no photons are detected, all M candidate symbols are equally likely. Since the received statistic takes binary values, the soft- and hard-decision capacities are equal.

When $K_b > 0$, we have $L(k) = e^{-K_s}\left(1 + [K_s/K_b]\right)^k$, and Eq. (4.6-2) becomes

$$C = \log_2(M)\left(1 - \frac{1}{\log_2 M} E_{Y_1,\cdots,Y_M} \log_2 \left[\sum_{i=1}^{M} \left(1 + \frac{K_s}{K_b}\right)^{(Y_j - Y_1)} \right]\right)$$

which is expressed in bits/slot. The case $K_b = 1$ is illustrated in Fig. 4-24 as a function of average power $P_{av} = K_s/M$ for a range of M. In the plot, an average

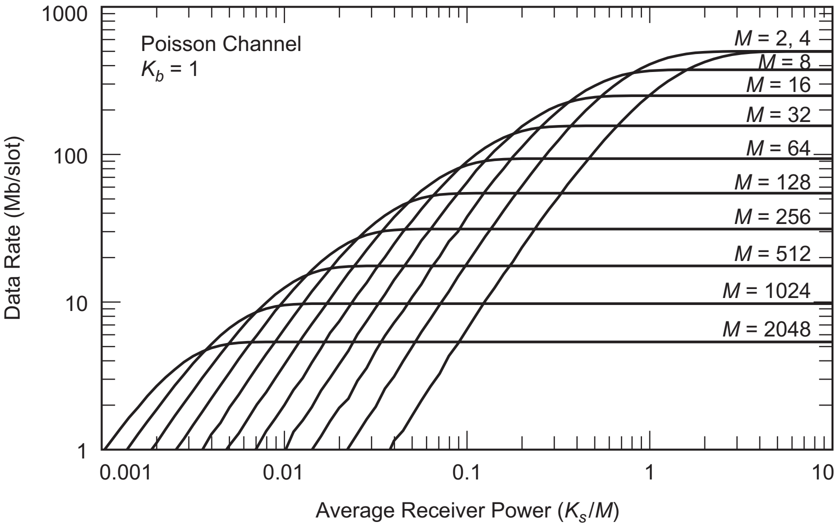

Fig. 4-24. Capacity of *M*-PPM on a Poisson channel, when K_b = 1.

power constraint would be represented by a vertical line. A peak constraint can be shown [77] to result in an upper limit on the PPM order. Hence, the maximum data rate subject to both peak and average power constraints can be identified using Fig. 4-24.

4.6.2.2 AWGN Channel. For the AWGN channel, the likelihood ratio reduces to

$$L(y) = \sqrt{\frac{\gamma}{\beta+\gamma}} \exp\left[\frac{\beta v^2 + 2\gamma\sqrt{\beta} v - \gamma\beta}{2(\beta+\gamma)}\right]$$

where $v = (y - m_0)/\sigma_0$ (recall β, γ are defined in Section 4.2.3). The capacity reduces to

$$C = \log_2 M - E \log_2 \sum_{j=1}^{M} \exp\left[\frac{(Y_j - Y_1)\big(\beta(Y_j + Y_1) + 2\gamma\sqrt{\beta}\big)}{2(\beta+\gamma)}\right]$$

bits per PPM symbol, or, when $\sigma_1^2 = \sigma_0^2$,

$$C = \log_2 M - E \log_2 \left[\sum_{j=1}^{M} e^{\sqrt{\beta}(Y_j - Y_1)}\right] \text{ bits/PPM symbol} \qquad (4.6\text{-}4)$$

4.6.3 Hard-Decision Versus Soft-Decision Capacity

The gap between hard- and soft-decision capacity for the optical channel is not fixed, but varies with the channel model and operating conditions. For example, the gap for the Poisson channel is zero when $K_b = 0$ (hard and soft decisions are equivalent) and increases to several dB with increasing K_b. In this section, we illustrate the gap for several channel models.

Figure 4-25 compares capacities for hard- and soft-decision AWGN channels for the case of $M = 256$. A similar comparison of capacities is shown in Fig. 4-26 for the hard-output and soft-output WMC channels. The capacity curves for both the AWGN and the WMC channels show that a minimum value of ρ_b is reached at a nonzero code rate. Unlike the soft-output channels, which exhibit monotonically better efficiency in terms of the bit-normalized SNR parameter ρ_b as the code rate (and hence the capacity per channel use) is reduced toward zero, the bit-normalized SNR efficiency of the hard-output channel worsens if the capacity is lowered below about 4 bits per channel use. This implies that an optimum code rate of about 1/2 will achieve the lowest ρ_b for the hard-output channel, while the soft-output channel achieves lowest ρ_b in the limit as the code rate goes to 0.

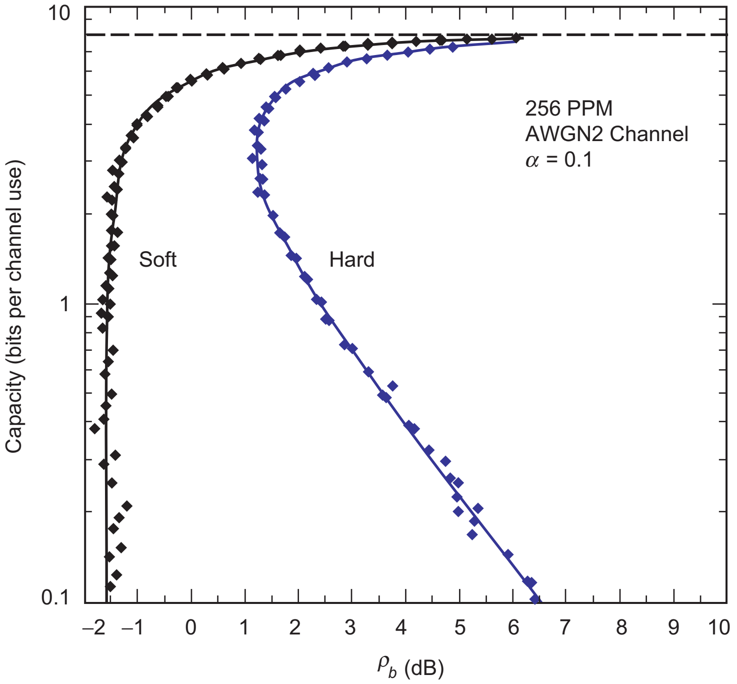

Fig. 4-25. Capacity of 256-PPM on hard-output and soft-output AWGN channels.

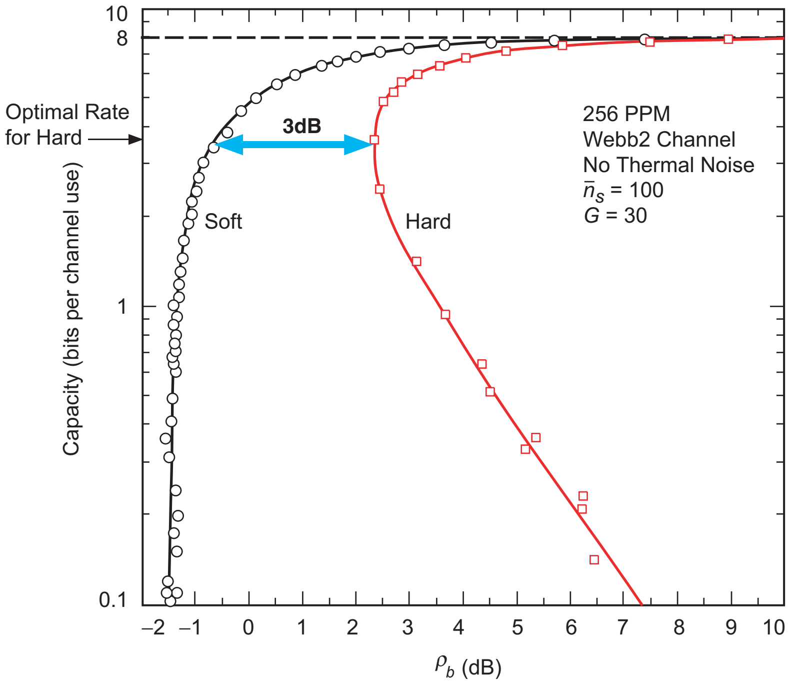

Fig. 4-26. Capacity of 256-PPM on hard-output and soft-output WMC channels.

Another comparison of capacity for the hard-output and soft-output WMC models is shown in Fig. 4-27, this time plotted versus $\bar{n}_b$. The hard-output capacity in this figure is based on Eq. (4.6-1) and was computed in [14] for a general WMC+Gaussian channel that also models the effects of thermal noise.

4.6.4 Losses Due to Using PPM

What loss is incurred by restricting the modulation to PPM? PPM is essentially a binary modulation code with a duty cycle $1/M$, and a single pulse (binary 1) in each (synchronized) window of M slots. Suppose we were to replace PPM with a binary modulation code with duty cycle $1/M$ but no constraint on the distribution of pulses. What gains are available by allowing an arbitrary pulse distribution?

The capacity of a memoryless channel with input restricted to duty cycle $1/M$ is

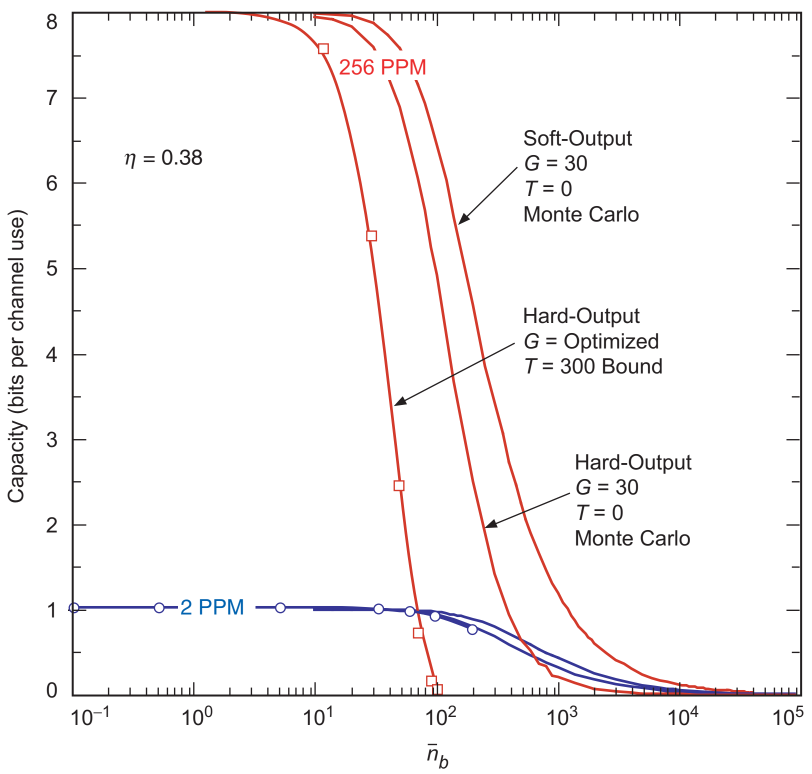

Fig 4-27. Comparison of capacity for hard-output and soft-output WMC model.

$$C_{\mathrm{OOK}} = \frac{1}{M} E_Y \log \frac{f_{Y|X}(Y|0)}{f_Y(Y)} + \frac{M-1}{M} E_Y \log \frac{f_{Y|X}(Y|1)}{f_Y(Y)} \tag{4.6-5}$$

where

$$f_Y(y) = \frac{1}{M} f_{Y|X}(y|0) + \frac{M-1}{M} f_{Y|X}(y|1)$$

is the probability mass function for a randomly chosen slot.

Let C^* be the capacity of the PPM channel optimized over the choice of PPM order (PPM orders are implicitly constrained to be powers of 2), and C^*_{OOK} the capacity of the duty-cycle-constrained channel optimized over the duty cycle (the duty cycle may take any positive real value). Figure 4-28 illustrates C^*_{OOK}/C^* for the Poisson channel as a function of the average power for a range of back-

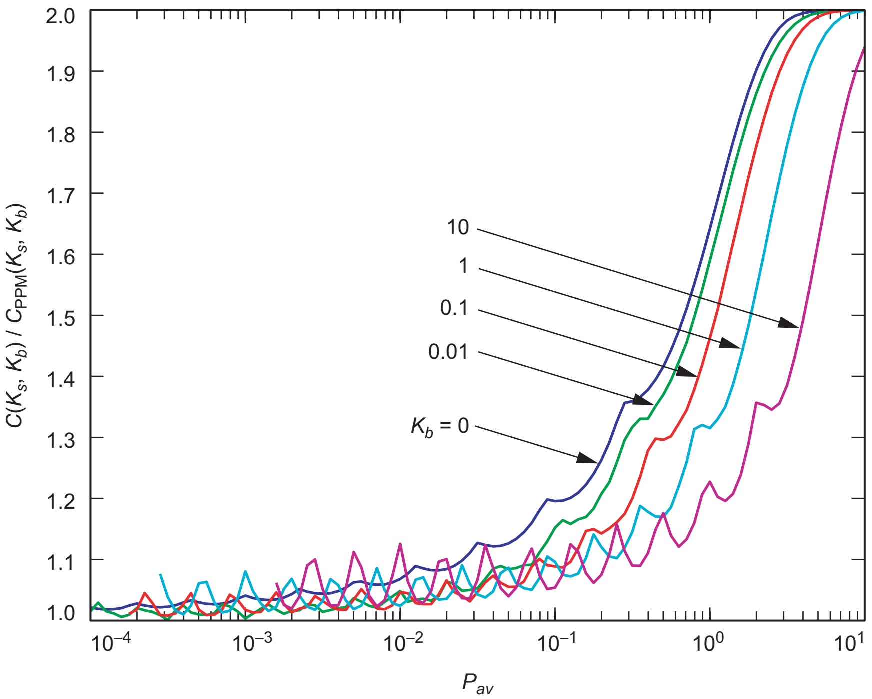

Fig. 4-28. Relative loss due to using PPM, C^*_{OOK}/C^*, Poisson channel, $K_b \in \{0, 0.01, 0.1, 1.0, 10\}$.

ground noise levels. This represents the potential gain in using an arbitrary duty-cycle constraint relative to PPM. The gains are larger for high average power, corresponding to small PPM orders, and for smaller background noise levels. We can potentially double the capacity for moderate to high average power. We note, however, in this discussion we have not specified codes that achieve arbitrary duty cycles. There are systematic methods to construct such codes, e.g., [78], but we will not explore their use here. We provide the results shown in Fig. 4-28 to demonstrate regions where their use should be explored. Since the deep-space optical channel is typically in the lower average power regime and losses in constraining the modulation to PPM are small in this area, in the remainder of the chapter we focus on results for PPM.

4.6.5 Capacity of the Binary Channel with Quantum Detection

The capacity of the binary OOK channel with quantum detection and with no external noise has been determined in [75]. Note that for the "noiseless" quantum model photon counting leads to an erasure channel whereas optimum quantum decoding results in a binary symmetric channel (BSC). For an arbitrary rotation of the measurement states with respect to the signal states, as described in [75], the transition probabilities are not equal, and, hence, a generalized (asymmetric) binary channel model must be considered. The capacity of the binary channel can be found by computing the mutual information between input and output for each rotation of the measurement states, starting with photon counting where one of the measurement states is aligned with the ground state, and computing the mutual information as a function of symbol input probability, β, for each rotation away from this configuration. For each rotation, the maximum of the mutual information as a function of β is recorded. The global maximum of the mutual information over all input probabilities and rotations is the capacity of the binary channel.

The input alphabet is denoted by A, and the output alphabet is denoted by B. The input alphabet consists of the two symbols $a_1 = 0$ and $a_2 = 1$. Likewise, the output can take on one of two values, namely $b_1 = 0$ and $b_2 = 1$. The probability that a 0 is transmitted is β, whereas the probability of a transmitted 1 is $1 - \beta \equiv \overline{\beta}$. The probability that b_2 is received (given that a_1 was transmitted) is p, while the probability that b_1 is received (given that a_2 was transmitted) is q.

The mutual information for this binary channel can be expressed as

$$\begin{aligned} I(A;B) = \; & \left[\left(\beta\bar{p} + \overline{\beta} q\right) \log\left(\frac{1}{\beta\bar{p} + \overline{\beta} q}\right) + \left(\beta p + \overline{\beta}\bar{q}\right) \log\left(\frac{1}{\beta p + \overline{\beta}\bar{q}}\right)\right] \\ & - \left[\beta\left(p \log\frac{1}{p} + \bar{p}\log\frac{1}{\overline{\beta}}\right) + \overline{\beta}\left(q\log\frac{1}{q} + \bar{q}\log\frac{1}{\bar{q}}\right)\right] \qquad (4.6\text{-}6) \end{aligned}$$

Note that the mutual information of the erasure and the BSC can be obtained by setting $p = 0$ and $p = q$, respectively.

The capacity of the quantum channel was determined by starting with a rotation angle of zero between the ground state and its measurement state (corresponding to photon counting, as we have shown above), and computing the mutual information defined in Eq. (4.6-6) as a function of β, $0 \le \beta \le 1$ for each rotation in the signal plane, until the measurement state corresponds to the signal state. Since different rotations yield different projections onto the measurement states, the values of p and q change with each rotation.

Examples of mutual information and capacity for the binary channel with OOK modulation are shown in Fig. 4-29, as a function of the input probability β,

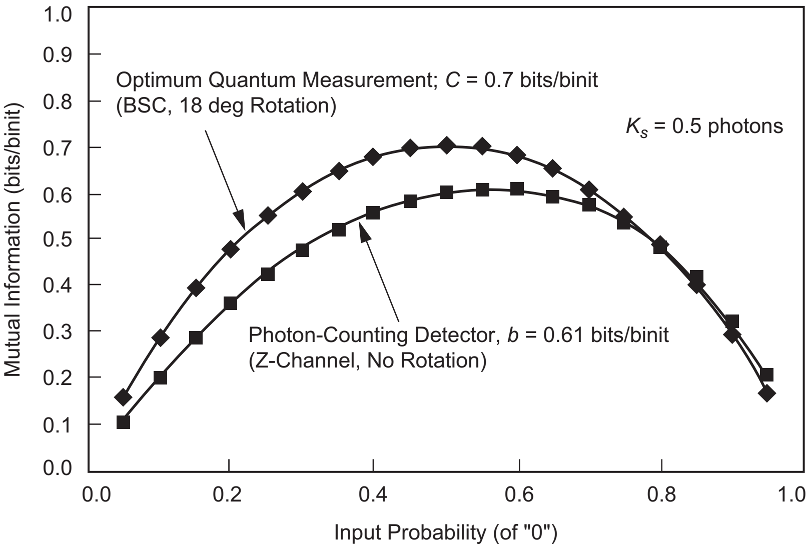

Fig. 4-29. Mutual information and capacity of the binary channel, with quantum and "classical" detection.

for an average value of one photon per symbol (or two photons per signal pulse). Only the limiting cases of optimum quantum measurement and photon counting are included. The error probabilities are approximately 0.1025 and 0.18 for quantum and direct detection, respectively. The global maximum value of mutual information was found to occur with optimum quantum measurement, at an input probability of $\beta = 0.5$. With photon-counting detection, for which the asymmetric z-channel is the correct representation, the maximum mutual information occurs at a higher value of input probability, namely at $\beta = 0.55$. The value of the maximum mutual information was found to be 0.7 bits/binit for quantum detection and 0.61 bits/binit for photon counting, verifying that optimum quantum detection achieves higher capacity.

4.7 Channel Codes for Optical Modulations

The constrained codes, or modulations, introduced in Section 4.3 enforce physical constraints and achieve desired peak-to-average power ratios, but their performance is far from the theoretical limits given in Section 4.6. In this section, we examine the application of Reed–Solomon, convolutional, turbo-like serial and parallel concatenated codes, and low-density parity check (LDPC) codes to the optical channel.

4.7.1 Reed–Solomon Codes

Since their introduction in 1960, Reed–Solomon (RS) codes [79] have become one of the most ubiquitous error-correcting codes. They have found applications in storage devices (tape, compact disk, digital video disc), bar codes, wireless communications (cellular telephones, deep-space RF communications), digital television, and high speed modems (digital subscriber line, DSL), in addition to optical communications.

An RS(n, k) code is a linear block code which encodes every block of k data symbols into n code symbols, where each symbol is an element of the Galois field with $q = n+1$ elements, denoted GF(q) [79]. Most commonly, q is a power of 2, $q = 2^s$, in which case each symbol is conveniently represented by s bits. Thus, the code can also be viewed as a (sn, sk) binary code.

In systematic form, a codeword of RS(n, k) contains k systematic (unchanged data) symbols and $n - k$ parity symbols. RS codes are maximum distance separable, meaning that they have the largest minimum distance, $d_{min} = n - k + 1$, among all linear (n, k) codes defined over GF(q). In some sense then, RS codes are optimal for their block length.[2] An RS(n, k) code can correct any pattern of $t = (d_{min} - 1)/2 = (n - k)/2$ errors; alternatively, it can correct any pattern of $2t$ erasures.

The encoder for RS codes operates as follows. The sk-bit message at the input to the RS encoder can be written as coefficients of a polynomial of degree $k - 1$:

$$u(X) = u_0 + u_1X + u_2X^2 + \cdots + u_{k-1}X^{k-1} \tag{4.7-1}$$

where each u_i is an element of GF(q), i.e., an s-bit block. An RS code has an associated generator polynomial

$$\begin{aligned} g(X) &= (X - \alpha)(X - \alpha^2)\cdots(X - \alpha^{n-k}) \\ &= g_0 + g_1X + g_2X^2 + \cdots + g_{n-k-1}X^{n-k-1} + X^{n-k} \end{aligned}$$

---

[2] However, an equivalent block-length code could have a higher minimum distance. For example, RS(255,223) has $d_{min} = 33$, the largest possible for this length code over GF(256). The same RS code could be viewed as a $(255 \times 8, 223 \times 8) = (2040, 1784)$ binary code, which is not necessarily a maximum distance separable code over GF(2), i.e., there could be other (2040,1784) binary codes with larger minimum distances.

where α is a primitive element of GF(q), and for each i, g_i is an element of GF(q). The parity symbols are the coefficients of the remainder

$$p(X) = p_0 + p_1 X + p_2 X^2 + \cdots + p_{n-k-1} X^{n-k-1} \tag{4.7-2}$$

that results from dividing $u(X)X^{n-k}$ by $g(X)$. There are efficient hardware implementations of this operation using shift registers [79]. To decode RS codes, $n-k$ syndromes are computed by dividing the received polynomial by $X + \alpha^i$, $i \in \{1, \cdots, n-k\}$. Then Berlekamp's algorithm is used to find the error-location polynomial, and the corrected values in the error-locations can be computed [79].

RS codes naturally fit the nonbinary nature of PPM signaling [21,30,31]. One can use RS(n, k) with M-PPM, $M = n+1$, by assigning each RS code symbol to one PPM symbol. When system constraints push one towards small M, this leads to small block length codes, which have limited coding gain, but this problem can be overcome by using a longer RS code and splitting RS code symbols across multiple PPM symbols.

If each of the n PPM symbols contains a received laser pulse in one of M slots, of average photon-energy K_s photons, the information rate for RS codewords is given by $\rho = ks/nK_s = r(s/K_s)$ bits/photon, where $r = k/n$ is the code rate.

4.7.2 Turbo and Turbo-Like Codes for Optical Modulations

For a fixed rate k/n, the performance of a code (e.g., the achievable bit error rate as a function of signal power), roughly speaking, increases with n, the block length of the code. However, the complexity of ML decoding of the code increases with the block length as well. For example, complexity increases quadratically in the block length with RS codes, making their implementation prohibitively complex for very large block lengths.

Turbo-like serial and parallel concatenated codes, decoded iteratively, achieve a large effective block length while providing low complexity near ML decoding. In this section, we discuss recent studies applying turbo and turbo-like codes to the deep-space, or free space, optical channel.

4.7.2.1 Parallel Concatenated (Turbo) Codes. A turbo code, or parallel concatenated convolutional code, preceding a PPM modulator has been proposed in [33,49]. The turbo code is a binary code, and its outputs are gathered and mapped to PPM symbols using the method described in Section 4.3.2. The turbo code can be viewed as an outer code C_o and the modulation as an inner code C_i.

To decode such a code effectively, the decoder needs more than the hard M-PPM symbol decisions that are used for RS-coded PPM. A turbo decoder takes from the receiver the set of M soft statistics corresponding to the slots, from which it can compute the likelihood of any symbol, and ultimately, the likelihood of each bit used to form the M-ary symbol. This is used to initialize the binary symbol likelihoods that a turbo decoding algorithm would use for conventional BPSK, as described in [80–82].

Turbo-coded PPM has been shown to offer improvements of 0.5 to 1 dB over RS-coded PPM, when the PPM size is 256 and WMC-plus-Gaussian statistics are assumed. Some additional improvement could be expected by updating the binary symbol likelihoods that are input to the turbo decoder with each new iteration. That is the approach taken in the next section, which also uses a single convolutional code instead of a parallel concatenation of convolutional codes.

4.7.2.2 Serially Concatenated Codes with Iterative Decoding. Modulation is a mapping of bits to symbols transmitted on the channel. This mapping may be considered a code and demodulation a decoding of the code. Conventionally, the modulation and ECC are decoded independently, with the demodulator sending its results to the ECC decoder. However, we may consider the combination of the modulation and the ECC as a single large code that maps user information bits directly to the symbols transmitted on the channel. We could gain several dB in performance by decoding the ECC and modulation jointly as a single code relative to decoding them independently.

An exact ML decoding of the joint modulation–ECC code would, in most cases of practical interest, be prohibitively complex. However, we may approximate true ML decoding while limiting the decoder complexity by iteratively decoding the modulation and the ECC. Iterative decoding is described in more detail in [80]. Applications of iterative decoding to the deep-space optical channel may be found in [84].

The PPM mapping may be preceded by a binary accumulator, making the inner code recursive. We refer to the inner code that is formed by the concatenation of a binary accumulator and PPM mapping as accumulate-PPM (APPM), illustrated in Fig. 4-30. The accumulator performs an exclusive-OR operation with the input bit and the previous output of the accumulator. This introduces memory in the modulator, which can be helpful in conjunction with an outer code.

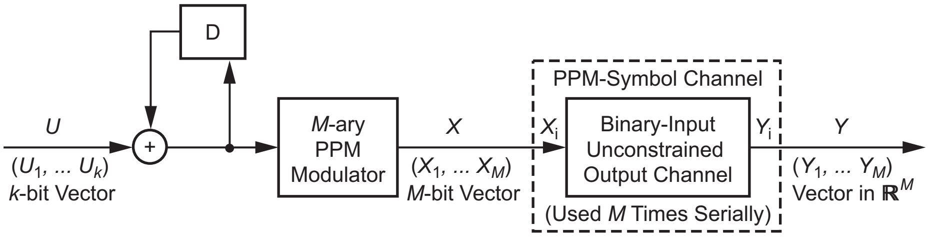

Fig. 4-30. APPM signaling.

4.8 Performance of Coded Optical Modulations

The prior sections have provided us with statistical models of the channel (Section 4.2), practical modulation formats for these channels that are physically realizable (Sections 4.3 and 4.4), the performance of the uncoded modulation (Section 4.5), the capacity of the channel when using this modulation (Section 4.6), and error correction codes suitable for concatenation with the modulation (Section 4.7). We now have all the pieces necessary to design a coded modulation for a particular channel and to measure its performance. In this section we provide general guidelines for this choice and illustrate performance measurements. Our running example will be the Poisson channel. The analysis would carry through in an analogous manner for other channel models.

4.8.1 Parameter Selection

In the absence of other constraints, we first choose to use the PPM order that maximizes the capacity for the available signal power. Other constraints are considered elsewhere [35–39,44–46]. For example, suppose our channel is Poisson with $K_b = 1$ and we have an average signal power of $K_s/M = 0.0541$ signal photons per slot. From Fig. 4-24, we find the the maximum capacity is 56 megabits (Mb) per slot, achieved by taking $M = 64$. Let M^* be the optimum choice of M. Figure 4-31 illustrates M^* constrained to be a power of two as a function of average power for the Poisson channel with $K_b \in \{0, 0.01, 0.1, 1.0, 10\}$. Discontinuities correspond to switching the order and are shown as vertical lines. The optimal order is increasing in K_b and decreasing in P_{av}, showing that as channel conditions become worse, the channel is used efficiently by moving to higher orders, increasing the peak power for a given average power. As channel conditions improve, we may increase the throughput by lowering the peak power and transmitting pulses more frequently (for the same average power).

Given the PPM order M^*, we choose the ECC code rate to satisfy

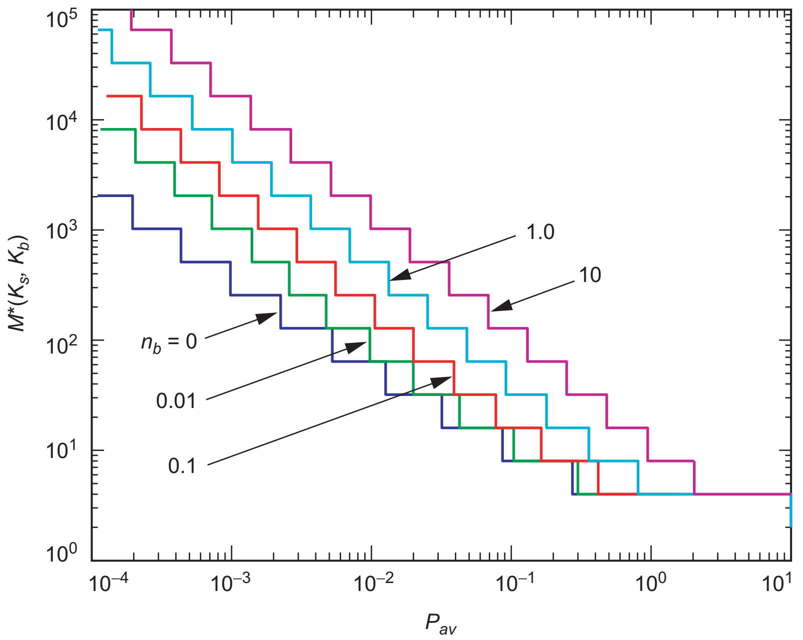

Fig. 4-31. Optimal PPM order, M^*, $K_b \in \{0, 0.01, 0.1, 1, 10\}$**.**

$$R \approx \frac{CM^*}{\log_2 M^*}$$

Figure 4-32 illustrates the ECC rate as a function of average power for several background levels on the Poisson channel. The discontinuities, corresponding to switching the order, obscure the general behavior. The behavior of the rate can be seen more clearly by allowing M^* to be real-valued. Let M^*_{OOK} specify the duty cycle of the modulation that maximizes Eq. (4.6-5). The upper bound on the rate of a modulation with duty cycle $1/M^*_{\text{OOK}}$ is given by

$$h\left(M^*_{\text{OOK}}\right) = \frac{1}{M^*_{\text{OOK}}} \log_2 M^*_{\text{OOK}} + \frac{M^*_{\text{OOK}} - 1}{M^*_{\text{OOK}}} \log_2 \left(\frac{M^*_{\text{OOK}}}{M^*_{\text{OOK}} - 1}\right)$$

bits per slot. The capacity and duty cycle may be used to specify an ECC data rate as

$$R_{\text{OOK}} = \frac{C^*_{\text{OOK}}}{h\left(M^*_{\text{OOK}}\right)}$$

which is shown in Fig. 4-33 for the Poisson channel with $K_b \in \{0, 0.01, 0.1, 1, 10\}$.

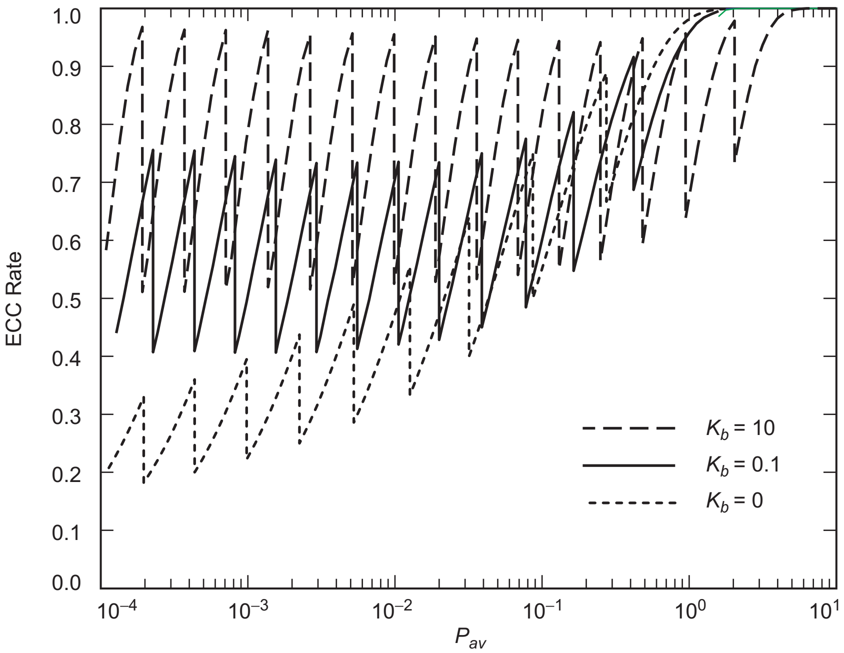

Fig. 4-32. PPM ECC rates, Poisson channel, $K_b \in \{0, 0.1, 10\}$.

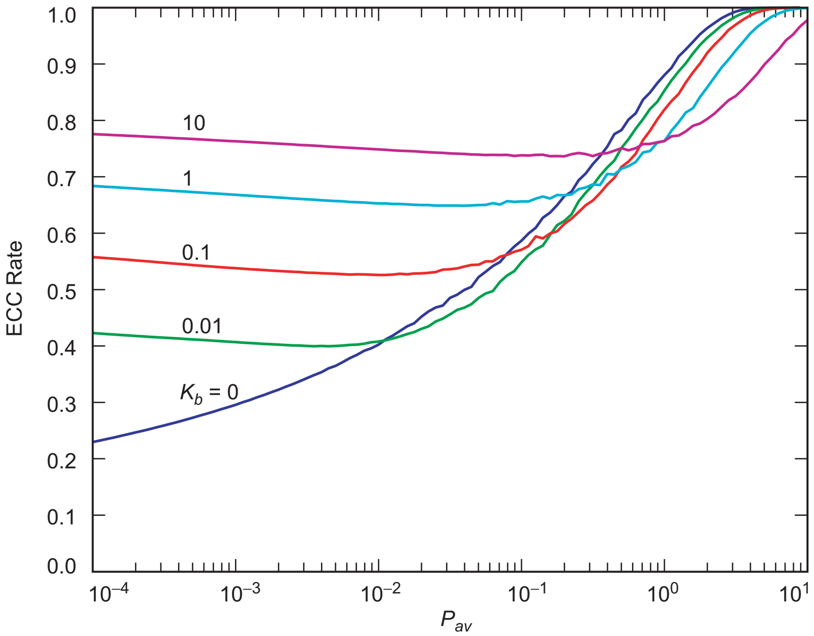

Fig. 4-33. Duty-cycle-constrained ECC rates, Poisson channel, $K_b \in \{0, 0.01, 0.1, 1, 10\}$.

4.8.2 Estimating Performance

We can save simulation time and gain insight by analytically determining the performance of coded modulation schemes. Such analysis is well known for RS codes and has been recently developed for iterative decoding schemes. An extensive discussion is beyond the scope of this chapter. We give a brief overview of the analysis in the following.

4.8.2.1 Reed–Solomon Codes. With hard-decision decoding, symbol decisions are made and sent to the RS decoder, which corrects all patterns of t or fewer errors. With p denoting the input channel symbol error probability, the probability of symbol error is approximated by the following expression:

$$P_s \approx (2^m - 1)^{-1} \sum_{k=t+1}^{N} k \binom{N}{k} p^k (1-p)^{N-k} \tag{4.8-1}$$

This expression holds for channel conditions dominated by symbol errors, rather than erasures. However, when channel conditions are dominated by symbol erasures, rather than symbol errors, the performance of the RS decoder improves.

An erasure RS decoder does not make undetected errors. That is, either the number of erasures in a received word is d_{min} or less, in which case the codeword is decoded correctly, or there are more than d_{min} erasures, in which case the decoder announces that the number of erasures is too large to decode properly.

The codeword error rate of an (n,k) RS code used with M-PPM on an erasure channel is (see, e.g., [21,79])

$$P_w = \sum_{j=d_{min}}^{n} \binom{n}{j} P_s^j (1-P_s)^{n-j} \tag{4.8-2}$$

where $d_{min} = n - k + 1$ is the minimum distance of the code, and P_s is the symbol erasure rate as discussed in Section 4.5.2.

When a decoder error occurs, with high probability a minimum distance codeword is chosen. Since code symbols are equally likely to be in error with an RS code, approximately $(n-k+1)/n$ information symbols are decoded incorrectly. On average, half the bits that map to a symbol will be in error, hence the bit error rate may be approximated by [20]

$$P_b \approx \frac{1}{2}\frac{n-k+1}{n}P_w \tag{4.8-3}$$

Performance results for RS-encoded PPM symbols on the erasure channel have been computed in [4] and reproduced in Fig. 4-34 for the case $M = 16, 32, 64$, and 128, with $k = M/2, N = M - 1$, and $d_{\min} = (M/2) - 1$.

Conventionally, RS decoders have operated on hard decisions from the receiver due to the complexity of computing maximum-likelihood estimates from soft decisions. However, recent results have demonstrated efficient decoding of RS codes using soft-decision inputs, improving their performance.

With soft-decision decoding, the ML decoder selects the RS codeword with the highest metric. For the case of Poisson detection, the optimum metric consists of the sum of slot counts corresponding to each codeword. For example, if a particular codeword consisted of a laser pulse in the first slot of every PPM symbol, then the metric corresponding to that codeword would be obtained by adding up the detected counts from the first slot of every symbol. Although exact error probabilities are difficult to compute, useful bounds on the probability of codeword error have been derived [83]. For the case of soft-decision decoding, an upper bound on the probability of codeword error (PWE) can be found by a direct application of the union bound. Since the premise of the union

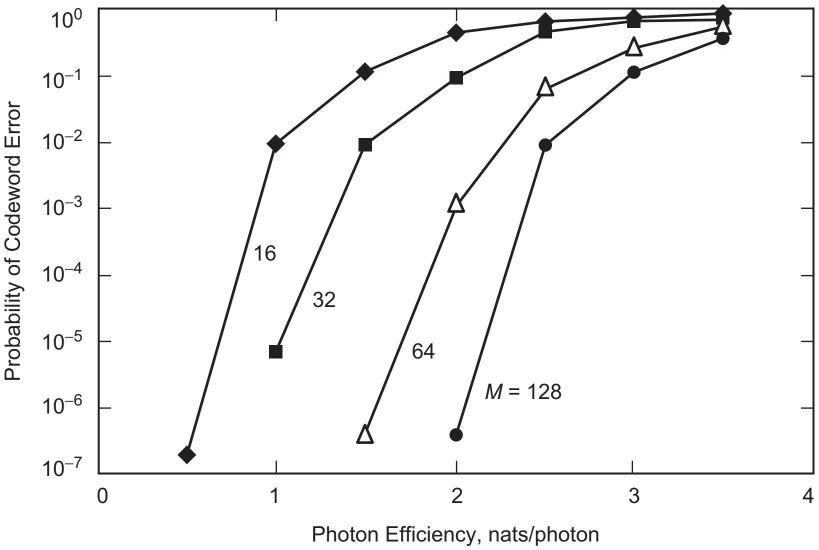

Fig. 4-34. RS decoding: codeword error probabilities for the erasure channel, M = 16, 32, 64, 128.

bound is that the exact error probability never exceeds the sum of pairwise error probabilities, we can write

$$P_w \leq \sum_{k=d_{\min}}^{n} L(k,n) P_k \tag{4.8-4}$$

where P_k is the probability of committing an error when attempting to decide between two codewords a distance k apart, and $L(k,n)$ is the codeword enumerator function that specifies the number of codewords that are a distance k apart, with a code of length n. For RS codes the enumerator function is of the form

$$L(k,n) = \sum_{j=n-k}^{n-d_{\min}} (-1)^{j+k-n} \binom{j}{N-k} \binom{n}{j} \left(M^{n-d_{\min}+1-j} - 1 \right) \tag{4.8-5}$$

where the index k takes on the values $k = d_{min}, d_{min}+1, \cdots n$. Plots of P_b as a function of n, for various average signal counts, and with $K_b = 1$ photon/slot, have been computed in [83].

4.8.2.2 Iterative Codes. The bit-error-rate versus average-power curve of a typical iterative code may be roughly divided into three regions (moving from left to right, or from low average power to high average power): a flat high error rate region, a "waterfall" region, and an "error floor" region. This behavior arises since typical iterative codes have a small minimum distance, hence an error floor, but low multiplicity, hence the waterfall region when (approximately) ML decoded. Recent results have illustrated methods to accurately predict the location of the waterfall and error floor.

The waterfall region may be predicted by SNR input–output diagrams, e.g., [85], or extrinsic information transfer (EXIT) curve charts, e.g., [86]. The error floor may be predicted from the first few terms of the weight enumerator polynomial and is ultimately dominated by the minimum distance and multiplicity of the minimum-weight codewords, e.g., [87].

4.8.3 Achievable Data Rates Versus Average Signal Power

In this section we present a sample code and modulation design, leading to a family of trade-offs of achievable data rate versus average signal power.

Suppose we have a system with a slot width of 1 ns and background noise $K_b = 1.0$. We would like to find the power required to achieve 56 megabits per second (Mbps) and choose appropriate coding and modulation. From Fig. 4-24, we find the optimum PPM order to achieve this data rate is $M = 64$, and the minimum required $K_s/M = 0.0541$. To achieve 56 Mbps, we choose a rate $R = 0.6 \approx 0.056/(\log_2(M)/M)$ ECC and concatenate it with 64-PPM. The performance of two candidate ECCs for this operating point, a serially concatenated convolutionally coded PPM, with outer code rate 3/5, and a 16,410-bit interleaver [SCPPM$(3/5, 2, 64, 16410)$] and Reed–Solomon coded PPM with $n = 4095$, $k = 2457$, and $M = 64$ [RSPPM$(4095, 2457, 64)$], are illustrated in Fig. 4-35.

Performance is compared with the lower bound on the bit error rate for a code with rate R, arrived at by application of the converse to the coding theorem [66, Theorem 7.2.2]:

$$P_b \geq \mathcal{H}^{-1}\left(1 - \frac{C}{R}\right) \tag{4.8-6}$$

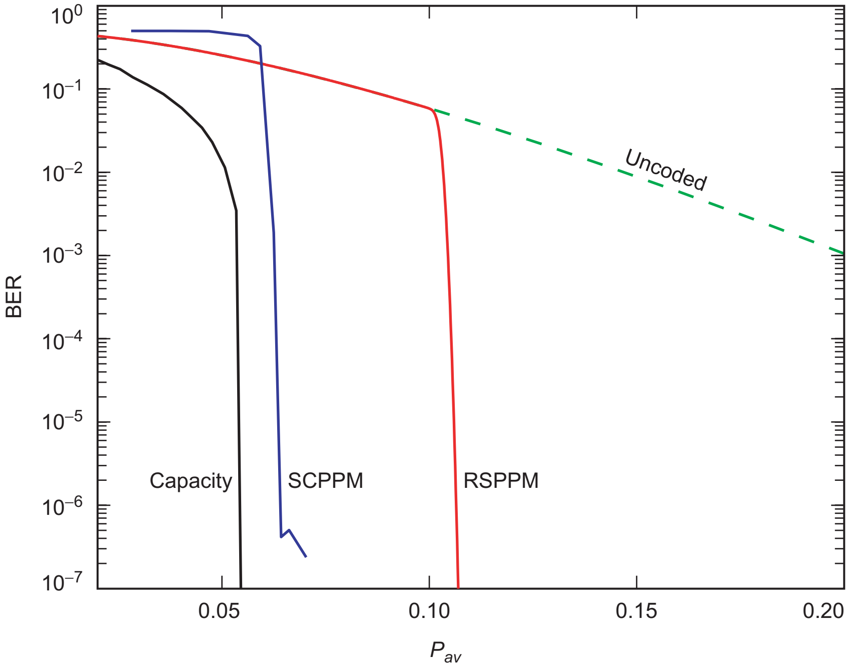

Fig. 4-35. Performance of SCPPM (3/5, 2, 64, 16410), RS (4095, 2457, 64), $K_b = 1.0$.

where $\mathcal{H}^{-1}$ is the inverse of the entropy function and C is the channel capacity. Unless otherwise noted, comparisons are made for bit error rates of 10^{-6}.

Figure 4-35 illustrates the bound in Eq. (4.8-6) for $C = 5.6$ Mbps as well as uncoded $M = 64$ performance, which, since it carries no coding redundancy, yields 9.4 Mbps. The SCPPM code operates 0.5 dB from capacity, the RS code operates 2.5–3 dB from capacity, and uncoded PPM operates 7.2 dB from capacity (at 5.6 Mbps). An appropriate comparison for uncoded 64-PPM is with capacity for 9.4 Mbps, from which uncoded performance is 4.7 dB. (It would be more efficient to achieve 9.4 Mbps with a rate 3/5 code mapped to 32-PPM).

These comparisons may be extended over a range of desired rates. Figure 4-36 illustrates achievable rates for $K_b = 1$ populated by points corresponding to the class of SCPPM codes, the class of RSPPM codes, and the uncoded PPM channel. The coded and uncoded channels are evaluated at a finite number of rates, which we connect in a line for illustration—this is justifiable by allowing time sharing. RSPPM points use the convention $n = M - 1$, although there is some degradation in using this convention for small M. For example, with $M = 64$, we illustrate a family of points where the RS code sym bol length is allowed to span multiple PPM symbols (as is done in Fig. 4-35).

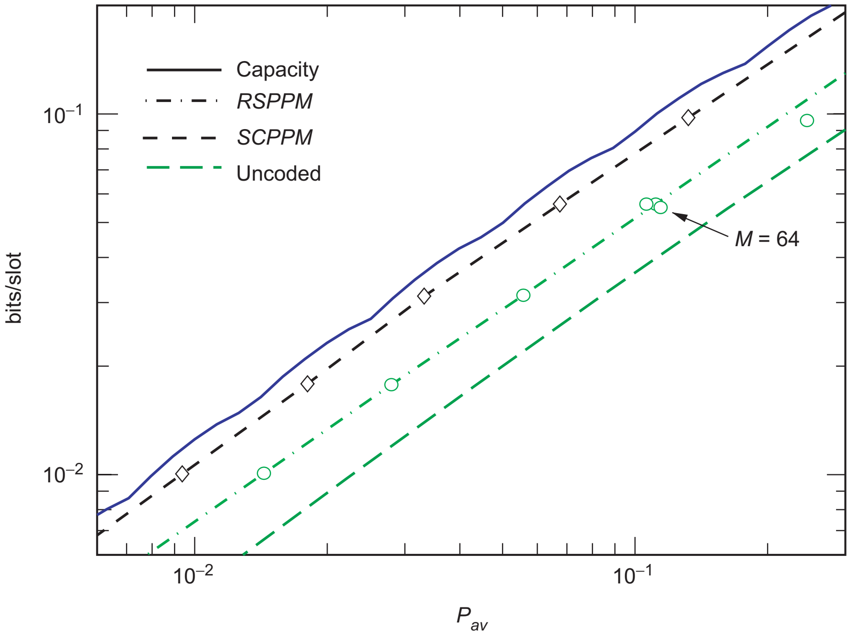

Fig. 4-36. Sample operating points, $K_b = 1$.

Points correspond to the average power at which the BER is 10^{-6}. We exclude iterative codes that exhibit error floors at BERs greater than 10^{-6}.

The class of SCPPM codes lies approximately 0.5 dB from capacity, while the class of RSPPM codes lies approximately 2.75 dB from capacity, and uncoded performance is 4.7 dB from capacity. These gaps will vary with K_b, but they provide a good approximation over a range of expected background noise levels.

References

[1] A. Schawlow and C. Townes, "Infrared and Optical Masers," *Phys. Rev.*, vol. 112, pp. 1940–1949, December 1958.

[2] T. H. Maiman, "Stimulated Optical Radiation in Ruby," *Nature*, vol. 187, p. 493, 1960.

[3] C. W. Helstrom, *Quantum Detection and Estimation Theory, Mathematics in Science and Engineering*, vol. 123, New York: Academic Press, 1976.

[4] R. J. McEliece and L. R. Welch, "Coding for Optical Channels with Photon-Counting," *The Deep Space Network Progress Report 42-52, May and June 1979*, Jet Propulsion Laboratory, Pasadena, California, pp. 61–66, August 15, 1979. http://ipnpr.jpl.nasa.gov/progress_report/

[5] C. W. Helstrom, J. W. S. Liu, and J. P. Gordon, "Quantum Mechanical Communications Theory," *Proceedings of the IEEE*, vol. 58, no. 10, pp. 1578–1598, 1970.

[6] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, July 1948.

[7] R. Gagliardi and S. Karp, *Optical Communications*, New York: J. Wiley and Sons, 1976.

[8] R. J. Glauber, "Coherent and Incoherent States of the Radiation Field," *The Physical Review*, vol. 131, pp. 2766–2788, September 1973.

[9] S. Barnett and P. Radmore, *Methods in Theoretical Quantum Optics*, New York: Oxford University Press, Inc., 1997.

[10] S. A. Butman, J. Katz, and J. R. Lesh, "Bandwidth Limitations on Noiseless Optical Channel Capacity," *IEEE Transactions on Communications*, vol. COM-30, pp. 1262–1264, May 1982.

[11] C.-C. Chen, "Figure of Merit for Direct-Detection Optical Channels," *The Telecommunications and Data Acquisition Progress Report 42-109, January–March 1992*, Jet Propulsion Laboratory, Pasadena, California, pp. 136–151, May 15, 1992. http://ipnpr.jpl.nasa.gov/progress_report/

[12] M. H. A. Davis, "Capacity and Cutoff Rate for Poisson-Type Channels," *IEEE Transactions on Information Theory*, vol. IT-26, pp. 710–715, November 1980.

[13] C. N. Georghiades, "Modulation and Coding for Throughput-Efficient Optical Systems," *IEEE Transactions on Information Theory*, vol. 40, pp. 1313–1326, September 1994.

[14] J. Hamkins, "The Capacity of Avalanche Photodiode-Detected Pulse-Position Modulation," *The Telecommunications and Mission Operations Progress Report 42-138, April–June 1999*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–19, August 15, 1999. http://ipnpr.jpl.nasa.gov/progress_report/

[15] C. W. Helstrom, "Comments on the Capacity of the Photon Counting Channel," *IEEE Transactions on Information Theory*, vol. IT-28, p. 556, May 1982.

[16] V. K. Jain, G. Steudal, and C. Rapp, "Channel Capacity for Optical Space Communication Systems," *Journal of Optical Communications*, vol. 18, no. 2, pp. 57–63, 1997.

[17] Y. M. Kabanov, "The Capacity of a Channel of the Poisson Type," *Theory Probab. Appl.*, vol. 23, pp. 143–147, 1978.

[18] J. R. Lesh, "Capacity Limit of the Noiseless, Energy-Efficient Optical PPM Channel," *IEEE Transactions on Communications*, vol. COM-31, pp. 546–548, April 1983.

[19] R. G. Lipes, "Pulse-Position-Modulation Coding as Near-Optimum Utilization of Photon Counting Channel with Bandwidth and Power Constraints," *The Deep Space Network Progress Report 42-56, January and February 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 108–113, April 15, 1980. http://ipnpr.jpl.nasa.gov/progress_report/

[20] J. L. Massey, "Capacity, Cutoff Rate, and Coding for a Direct-Detection Optical Channel," *IEEE Transactions on Communications*, vol. COM-29, pp. 1615–1621, November 1981.

[21] R. J. McEliece, "Practical Codes for Photon Communication," *IEEE Transactions on Information Theory*, vol. IT-27, pp. 393–398, July 1981.

[22] J. R. Pierce, E. C. Posner, and E. R. Rodemich, "The Capacity of the Photon Counting Channel," *IEEE Transactions on Information Theory*, vol. IT-27, pp. 61–77, January 1981.

[23] D. L. Snyder and I. B. Rhoades, "Some Implications of the Cutoff-Rate Criterion for Coded Direct-Detection Optical Communication Systems," *IEEE Transactions on Information Theory*, vol. IT-26, pp. 327–338, May 1980

[24] H. H. Tan, "Capacity of a Multimode Direct Detection Optical Communication Channel," *The Telecommunications and Data Acquisition Progress Report 42-63, March and April 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 51–70, June 15, 1981. http://ipnpr.jpl.nasa.gov/progress_report/

[25] M. Takahsi, H. Yashima, I. Sasase, and S. Mori, "Capacity and Effects of Reed–Solomon Codes on Multi-Pulse PPM in Optical Communications," *International Conference on Communication*, vol. 345.4.1, pp. 1663–1667, April 1990.

[26] A. D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel—Part I," *IEEE Transactions on Information Theory*, vol. 34, pp. 1449–1461, November 1988.

[27] A. D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel—Part II," *IEEE Transactions on Information Theory*, vol. 34, pp. 1462–1471, November 1988.

[28] D. Zwillinger, "Differential PPM has a Higher Throughput than PPM for the Band-Limited and Average-Power-Limited Optical Channel," *IEEE Transactions on Information Theory*, vol. 34, pp. 1269–1273, September 1988.

[29] A. R. Calderbank and C. N. Georghiades, "Synchronizable Codes for the Optical OPPM Channel," *IEEE Transactions on Information Theory*, vol. 40, pp. 1097–1107, July 1994.

[30] D. Divsalar, R. M. Gagliardi, and J. H. Yuen, "PPM Performance for Reed–Solomon Decoding over an Optical–RF Relay Link," *IEEE Transactions on Communications*, vol. COM-32, pp. 302–305, March 1984.

[31] E. Forestieri, R. Gangopadhyay, and G. Prati, "Performance of Convolutional Codes in a Direct-Detection Optical PPM Channel," *IEEE Transactions on Communications*, vol. 37, no. 12, pp. 1303–1317, 1989.

[32] R. M. Gagliardi and S. Karp, *Optical Communication*, New York: John Wiley and Sons, Inc., 1995.

[33] J. Hamkins, "Performance of Binary Turbo-Coded 256-ary Pulse-Position Modulation," *The Telecommunications and Mission Operations Progress Report 42-138, AprilJune 1999*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–15, August 15, 1999. http://ipnpr.jpl.nasa.gov/progress_report/

[34] M. Jeganathan and S. Mecherle, *A Technical Manual for FOCAS 2.0—Free-Space Optical Communications Analysis Software*, May 1998.

[35] C.-H. Lai and K. Kiasaleh, "Modified Viterbi Decoders for Joint Data Detection and Timing Recovery of Convolutionally Encoded PPM and OPPM Optical Signals," *IEEE Transactions on Communications*, vol. 45, pp. 90–94, January 1997.

[36] C.-H. Lai and K. Kiasaleh, "A New Performance Upper Bound for Convolutionally Encoded Direct-Detection Optical OPPM Communications Systems," *Proceedings of the International Conference on Communications (ICC'95)*, Seattle, Washington, pp. 1302–1306, June 1995.

[37] V. Vilnrotter, *Optical Receivers Using Rough Reflectors*, JPL Publication 25, Jet Propulsion Laboratory, Pasadena, California, May 15, 1985.

[38] V. A. Vilnrotter, "Spatial Acquisition of Optical Sources in the Presence of Intense Interference," *The Telecommunications and Data Acquisition Progress Report 42-58, May and June 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 91–96, August 15, 1980.
http://ipnpr.jpl.nasa.gov/progress_report/

[39] Z. Xiaomai, X. Zhaofei, and H. Tiexin, "On Evaluating the Performance of Error-Correcting Codes in Lasercom PPM Systems with Direct Detection," *IEEE TENCON*, Beijing, China, pp. 489–491, 1993.

[40] R. J. McIntyre, "The Distribution of Gains in Uniformly Multiplying Avalanche Photodiodes: Theory," *IEEE Transactions on Electronic Devices*, vol. ED-19, no. 6, pp. 703–713, June 1972.

[41] P. P. Webb, R. J. McIntyre, and J. Conradi, "Properties of Avalanche Photodiodes," *RCA Review*, vol. 35, pp. 234–278, June 1974.

[42] F. M. Davidson and X. Sun, "Gaussian Approximation Versus Nearly Exact Performance Analysis of Optical Communication Systems with PPM Signaling and APD Receivers," *IEEE Transactions on Communications*, vol. 36, pp. 1185–1192, November 1988.

[43] S. J. MacMullan and O. M. Collins, "The Capacity of Orthogonal and Bi-Orthogonal Codes on the Gaussian Channel," *IEEE Transactions on Information Theory*, vol. 44, pp. 1217–1232, May 1998.

[44] R. A. Cryan, A. J. Phillips, and J. M. Senior, "Optically Preamplified n-ary PPM Systems," *SPIE*, vol. 2024, pp. 72–80, 1993.

[45] T. T. Ha, G. E. Keiser, and R. L. Borchardt, "Analysis of Direct Detection Lightwave Systems with Optical Amplifiers," *Proceedings of MILCOM (Military Communications Conference)*, pp. 14.4-1–14.4-5, 1994.

[46] M. A. Herro and L. Hu, "A New Look at Coding for APD-Based Direct-Detection Optical Channels," *IEEE Transactions on Information Theory*, vol. 34, pp. 858–866, July 1988.

[47] L. W. Hughes, "A Simple Upper Bound on the Error Probability for Orthogonal Signals in White Noise," *IEEE Transactions on Communications*, vol. 40, p. 670, April 1992.

[48] G. S. Mecherle, *Maximized Data Rate Capability for Optical Communication Using Semiconductor Devices with Pulse Position Modulation*, Ph.D. Thesis, University of Southern California, Los Angeles, California, May 1986.

[49] J. Hamkins and M. Srinivasan, "Turbo Codes for APD-Detected PPM," *Proceedings of the Thirty-Sixth Annual Allerton Conference on Communication, Control, and Computing*, pp. 29–38, September 1998.

[50] M. Srinivasan and V. Vilnrotter, "Performance of the Optimum Receiver for Pulse-Position Modulation Signals with Avalanche Photodiode Statistics," *The Telecommunications and Mission Operations Progress Report 42-133, January–March 1998*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–10, May 15, 1998. http://ipnpr.jpl.nasa.gov/progress_report/

[51] S. Dolinar, D. Divsalar, J. Hamkins, and F. Pollara, "Capacity of Pulse-Position Modulation (PPM) on Gaussian and Webb Channels," *The Telecommunications and Mission Operations Progress Report 42-142, April–June 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–31, August 15, 2000. http://ipnpr.jpl.nasa.gov/progress_report/

[52] M. Srinivasan and V. Vilnrotter, "Symbol-Error Probabilities for Pulse-Position Modulation Signaling with an Avalanche Photodiode Receiver and Gaussian Thermal Noise," *The Telecommunications and Mission Operations Progress Report 42-134, April–June 1998*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–11, August 15, 1998. http://ipnpr.jpl.nasa.gov/progress_report/

[53] G. G. Ortiz, J. V. Sandusky, and A. Biswas, "Design of an Opto-Electronic Receiver for Deep-Space Optical Communications," *The Telecommunications and Mission Operations Progress Report 42-142, April–June 2000*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–17, August 15, 2000.
http://ipnpr.jpl.nasa.gov/progress_report/

[54] M. Srinivasan, J. Hamkins, B. Madden-Woods, A. Biswas, and J. Beebe, "Laboratory Characterization of Silicon Avalanche Photodiodes (APDs) for Pulse-Position Modulation (PPM) Detection," *The InterPlanetary Network Progress Report 42-146, April–June 2001*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–14, August 15, 2001.
http://ipnpr.jpl.nasa.gov/progress_report/

[55] G. M. Lee and G. W. Schroeder, "Optical PPM with Multiple Positions per Pulsewidth," *IEEE Transactions on Communications*, vol. COM-25, pp. 360–364, March 1977.

[56] R. A. Horn and C. R. Johnson, *Matrix Analysis*, New Jersey: Cambridge University Press, 1988.

[57] C. E. Shannon, *The Mathematical Theory of Communication*, Urbana, Illinois: University of Illinois Press, 1963.

[58] K. A. S. Immink, *Codes for Mass Data Storage Systems*, The Netherlands: Shannon Foundation Publishers, 1999.

[59] K. A. S. Immink, P. H. Siegel, and J. K. Wolf, "Codes for Digital Recorders," *IEEE Transactions on Information Theory*, vol. 44, pp. 2260–2299, October 1998.

[60] B. H. Marcus, R. M. Roth, and P. H. Siegel, "Constrained Systems and Coding for Recording Channels," *Handbook of Coding Theory*, V. S. Pless and W. C. Huffman, eds., Chapter 20, Elsevier Science, 1998.

[61] P. A. Franaszek, "Sequence-State Coding for Digital Transmission," *The Bell System Technical Journal*, vol. 47, pp. 143–157, 1968.

[62] H. V. Poor, *An Introduction to Signal Detection and Estimation*, New York,: Springer-Verlag, 1988.

[63] V. Vilnrotter and E. Rodemich, "A Generalization of the Near-Optimum Binary Coherent State Receiver Concept," *IEEE Transactions on Information Theory*, vol. IT-30, pp. 446–450, March 1984.

[64] R. M. Gagliardi, V. A. Vilnrotter, and S. J. Dolinar, *Optical Deep Space Communication via Relay Satellite*, JPL Publication 81-40, Jet Propulsion Laboratory, Pasadena, California, August 15, 1981.

[65] V. Vilnrotter, M. Simon, and M. Srinivasan, *Maximum Likelihood Detection of PPM Signals Governed by an Arbitrary Point Process Plus Additive Gaussian Noise*, JPL Publication 98-7, Jet Propulsion Laboratory, Pasadena, California, April 1998.

[66] R. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[67] J. Hamkins, "Accurate Computation of the Performance of M-ary Orthogonal Signaling on a Discrete Memoryless Channel," *IEEE Transactions on Communications*, 2003 (to appear).

[68] V. A. Vilnrotter, "An M-ary Coherent Optical Receiver for the Free-Space Channel," *The Telecommunications and Data Acquisition Progress Report 42-66, September and October 1981*, Jet Propulsion Laboratory, Pasadena, California, pp. 60–66, December 15, 1981.
http://ipnpr.jpl.nasa.gov/progress_report/

[69] R. S. Kennedy, "A Near-Optimum Receiver for the Binary Coherent State Quantum Channel," *M.I.T. Research Laboratory of Electronics Quarterly Progress Report*, vol. 108, pp. 219–225, January 1973.

[70] S. Dolinar, "An Optimum Receiver for the Binary Coherent State Quantum Channel," *M.I.T. Research Laboratory of Electronics Quarterly Progress Report*, vol. 111, pp. 115–120, October 1973.

[71] S. J. Dolinar, Jr. "A Near-Optimum Receiver Structure for the Detection of M-ary Optical PPM Signals," *The Telecommunications and Data Acquisition Progress Report 42-72, October–December 1982*, Jet Propulsion Laboratory, Pasadena, California, pp. 30–42, February 15, 1983.
http://ipnpr.jpl.nasa.gov/progress_report/

[72] V. A. Vilnrotter, "A Binary Coherent Optical Receiver for the Free-Space Channel," *The Telecommunications and Data Acquisition Progress Report 42-61, November and December 1980*, Jet Propulsion Laboratory, Pasadena, California, pp. 27–38, February 15, 1981.
http://ipnpr.jpl.nasa.gov/progress_report/

[73] V. Vilnrotter and C.-W. Lau, "Quantum Detection in the Presence of Noise," *Proceedings of SPIE, Free-Space Laser Communications Technologies XV*, vol. 4975, San Jose, California, January 2003.

[74] V. A. Vilnrotter and C.-W. Lau, "Quantum Detection of Binary and Ternary Signals in the Presence of Thermal Noise Fields," *The Interplanetary Network Progress Report 42-152, October–December 2002*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–13, February 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[75] V. Vilnrotter and C.-W. Lau, "Quantum Detection Theory for the Free-Space Channel," *The InterPlanetary Network Progress Report 42-146, April–June 2001*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–34, August 15, 2001.

[76] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.

[77] B. Moision and J. Hamkins, "Deep-Space Optical Communications Downlink Budget: Modulation and Coding," *The Interplanetary Network Progress Report 42-154, April–June 2003*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–28, August 15, 2003. http://ipnpr.jpl.nasa.gov/progress_report/

[78] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-State Modulation Codes for Data Storage," *IEEE Journal Selected Areas Communications*, vol. 10, pp. 5–37, January 1992.

[79] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications*, New Jersey: Prentice Hall, 1983.

[80] S. Benedetto, D. Divsalar, G. Montorsi, and F.Pollara, "A Soft-Input Soft-Output Maximum A Posteriori (MAP) Module to Decode Parallel and Serial Concatenated Codes," *The Telecommunications and Data Acquisition Progress Report 42-127, July–September 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–20, November 15, 1996. http://ipnpr.jpl.nasa.gov/progress_report/

[81] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[82] K. Andrews, V. Stanton, S. Dolinar, V. Chen, J. Berner, and F. Pollara, "Turbo-Decoder Implementation for the Deep Space Network," *The Interplanetary Network Progress Report 42–148, October–December 2001*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–20, February 15, 2002. http://ipnpr.jpl.nasa.gov/progress_report/

[83] G. Prati and R. Gagliardi, "Block Encoding and Decoding for the Optical PPM Channel," *IEEE Transactions on Information Theory*, vol. IT-28, January 1981.

[84] B. Moision and J. Hamkins, "Low Complexity Serially Concatenated Codes for the Deep-Space Optical Channel," International Symposium on Information Theory, Yokohama, Japan, June 2003.

[85] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative Turbo Decoder Analysis Based on Density Evolution," *IEEE Journal Selected Areas Communications*, vol. 19, pp. 891–907, May 2001.

[86] A. Ashikhmin, G. Kramer, and S. ten Brink, "Code Rate and the Area Under the Extrinsic Information Transfer Curves," *IEEE International Symposium on Information Theory*, p. 115, 2002.

[87] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial Concatenation of Interleaved Codes: Performance Analysis, Design, and Iterative Decoding," *The Telecommunications and Data Acquisition Progress Report 42-126, April–June 1996*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–26, August 15, 1996.
http://tmo.jpl.nasa.gov/tmo/progress_report/42-126/126D.pdf

Notation

Systems Parameters

$R =$ Data rate, bits/second

$\rho =$ Photon efficiency, bits/photon

$n_s =$ Rate of photons incident on detector, photons/second)

$P_b =$ Probability of bit error (BER)

$P_s =$ Probability of symbol error (SER)

$P_c =$ Probability of correct detection

$X =$ Value (typically binary) transmitted during a time slot

$Y =$ Real value at output of detector

$m, \sigma^2 =$ Mean, variance of Y

$f_{Y|X}(y|0) =$ Conditional probability density (or mass) function of Y given $X = 0$

$f_{Y|X}(y|1) =$ Conditional probability density (or mass) function of Y given $X = 1$

$m_0, m_1 =$ Conditional mean of Y given $X = 0$ or 1

$\sigma_0^2, \sigma_1^2 =$ Conditional variance of Y given $X = 0$ or 1

$K_s =$ Average number of absorbed signal photons per pulse

$K_b =$ Average number of absorbed background photons per slot

$\beta =$ Slot SNR, $(m_1 - m_0)^2/\sigma^2$

$\gamma =$ Excess slot SNR, $(m_1 - m_0)^2/\sigma_1^2 - \sigma_0^2$

$\beta_b =$ Bit SNR, $\beta/(2R_c)$

Laser and Modulator Parameters

$\nu =$ Optical frequency, nanometers

$M =$ PPM order

$T_s =$ Width of the signal slot, seconds

Detector Parameters

$\eta =$ Quantum efficiency

$k_{eff} =$ Ionization ratio

$F =$ Excess noise factor, $k_{eff}G + (2 - 1/G)(1 - k_{eff})$

$I_b =$ Bulk leakage current, amperes

$I_s =$ Surface leakage current, amperes

$T =$ Noise temperature, kelvins

$R_L =$ Load resistance (transimpedance model), $5.75 \times 10^{12} \times T_s$ ohms

$B =$ Noise equivalent one-sided bandwidth, hertz

$G =$ Gain

Physical Constants

$\kappa =$ Boltzmann's constant, 1.38×10^{-23} joules/kelvin

$q =$ Electron charge, 1.6×10^{-19} coulombs

$h =$ Planck's constant, 6.624×10^{-34} joules/hertz